



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/49, A61K 48/00	A2	(11) International Publication Number: WO 00/39302 (43) International Publication Date: 6 July 2000 (06.07.00)
(21) International Application Number: PCT/US99/31245 (22) International Filing Date: 30 December 1999 (30.12.99) (30) Priority Data: 60/114,495 31 December 1998 (31.12.98) US 60/168,471 1 December 1999 (01.12.99) US (71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).	(72) Inventors: BARNETT, Susan; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). ZUR MEGEDE, Jan; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). SRIVASTAVA, Indresh; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). LIAN, Ying; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). HARTOG, Karin; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). LIU, Hong; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). GREER, Catherine; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). SELBY, Mark; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). WALKER, Christopher; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). (74) Agents: DOLLARD, Anne, S.; Chiron Corporation, Intellectual Property - R440, P.O. Box 8097, Emeryville, CA 94662-8097 (US) et al. (81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(54) Title: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION OF VIRUS-LIKE PARTICLES		
(57) Abstract <p>The present invention relates to the efficient expression of HIV polypeptides in a variety of cell types, including, but not limited to, mammalian, insect, and plant cells. Synthetic expression cassettes encoding the HIV Gag-containing polypeptides are described, as are uses of the expression cassettes in applications including DNA immunization, generation of packaging cell lines, and production of Env-, tat- or Gag-containing proteins. The invention provides methods of producing Virus-Like Particles (VLPs), as well as, uses of the VLPs including, but not limited to, vehicles for the presentation of antigens and stimulation of immune response in subjects to whom the VLPs are administered.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND
PRODUCTION OF VIRUS-LIKE PARTICLES

5 TECHNICAL FIELD

Synthetic expression cassettes encoding the HIV polypeptides (e.g., Gag-, pol-, prot-, reverse transcriptase, Env- or tat-containing polypeptides) are described, as are uses of the expression cassettes. The present invention relates to the efficient expression of HIV polypeptides in a variety of cell types. Further, the invention provides methods of producing Virus-Like Particles (VLPs), as well as, uses of the VLPs and high level expression of oligomeric envelope proteins.

15

BACKGROUND OF THE INVENTION

Acquired immune deficiency syndrome (AIDS) is recognized as one of the greatest health threats facing modern medicine. There is, as yet, no cure for this disease.

20

In 1983-1984, three groups independently identified the suspected etiological agent of AIDS. See, e.g., Barre-Sinoussi et al. (1983) Science 220:868-871; Montagnier et al., in Human T-Cell Leukemia Viruses (Gallo, Essex & Gross, eds., 1984); Vilmer et al. (1984) The Lancet 1:753; Popovic et al. (1984) Science 224:497-500; Levy et al. (1984) Science 225:840-842. These isolates were variously called lymphadenopathy-associated virus (LAV), human T-cell lymphotropic virus

type III (HTLV-III), or AIDS-associated retrovirus (ARV). All of these isolates are strains of the same virus, and were later collectively named Human Immunodeficiency Virus (HIV). With the isolation of a related

- 5 AIDS-causing virus, the strains originally called HIV are now termed HIV-1 and the related virus is called HIV-2. See, e.g., Guyader et al. (1987) *Nature* 326:662-669; Brun-Vezinet et al. (1986) *Science* 233:343-346; Clavel et al. (1986) *Nature* 324:691-695.

- 10 A great deal of information has been gathered about the HIV virus, however, to date an effective vaccine has not been identified. Several targets for vaccine development have been examined including the *env*, *Gag*, *pol* and *tat* gene products encoded by HIV.

- 15 Haas, et al., (*Current Biology* 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (*J. Virol.* 72(2):1497-1503, 1998) described an increased immune
20 response elicited by DNA vaccination employing a synthetic gp120 sequence with optimized codon usage. Schneider, et al., (*J Virol.* 71(7):4892-4903, 1997) discuss inactivation of inhibitory (or instability) elements (INS) located within the coding sequences of the
25 *Gag* and *Gag-protease* coding sequences.

- The *Gag* proteins of HIV-1 are necessary for the assembly of virus-like particles. HIV-1 *Gag* proteins are involved in many stages of the life cycle of the virus including, assembly, virion maturation after particle
30 release, and early post-entry steps in virus replication. The roles of HIV-1 *Gag* proteins are numerous and complex (Freed, E.O., *Virology* 251:1-15, 1998).

Wolf, et al., (PCT International Application, WO 96/30523, published 3 October 1996; European Patent Application, Publication No. 0 449 116 A1, published 2 October 1991) have described the use of altered pr55 Gag of HIV-1 to act as a non-infectious retroviral-like particulate carrier, in particular, for the presentation of immunologically important epitopes. Wang, et al., (*Virology* 200:524-534, 1994) describe a system to study assembly of HIV Gag- β -galactosidase fusion proteins into virions. They describe the construction of sequences encoding HIV Gag- β -galactosidase fusion proteins, the expression of such sequences in the presence of HIV Gag proteins, and assembly of these proteins into virus particles.

Recently, Shiver, et al., (PCT International Application, WO 98/34640, published 13 August 1998) described altering HIV-1 (CAM1) Gag coding sequences to produce synthetic DNA molecules encoding HIV Gag and modifications of HIV Gag. The codons of the synthetic molecules were codons preferred by a projected host cell.

The envelope protein of HIV-1 is a glycoprotein of about 160 kD (gp160). During virus infection of the host cell, gp160 is cleaved by host cell proteases to form gp120 and the integral membrane protein, gp41. The gp41 portion is anchored in (and spans) the membrane bilayer of virion, while the gp120 segment protrudes into the surrounding environment. As there is no covalent attachment between gp120 and gp41, free gp120 is released from the surface of virions and infected cells.

Haas, et al., (*Current Biology* 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (*J. Virol.*

72(2):1497-1503, 1998) described an increased immune response elicited by DNA vaccination employing a synthetic gp120 sequence with optimized codon usage.

5 **SUMMARY OF THE INVENTION**

The present invention relates to improved expression of HIV *Env*-, *tat*-, *pol*-, *prot*-, *reverse transcriptase*, or *Gag*-containing polypeptides and production of virus-like particles.

10 In one embodiment the present invention includes an expression cassette, comprising a polynucleotide encoding an HIV *Gag* polypeptide comprising a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20. In certain embodiments, the polynucleotide
15 sequence encoding said *Gag* polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:9 or SEQ ID NO:4. The expression cassettes may further include a polynucleotide sequence encoding an HIV *protease* polypeptide, for
20 example a nucleotide sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:5, SEQ ID NO:78, and SEQ ID NO:79. The expression cassettes may further include a polynucleotide sequence encoding an HIV *reverse*
25 *transcriptase* polypeptide, for example a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, and SEQ ID NO:84. The expression cassettes may further include a polynucleotide
30 sequence encoding an HIV *tat* polypeptide, for example a sequence selected from the group consisting of: SEQ ID NO:87, SEQ ID NO:88, and SEQ ID NO:89. The expression cassettes may further include a polynucleotide sequence encoding an HIV *polymerase* polypeptide, for example a

sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6. The expression cassettes may include a polynucleotide sequence encoding an HIV *polymerase* polypeptide, wherein (i) the nucleotide
5 sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:4, and (ii) wherein the sequence is modified by deletions of coding regions corresponding to reverse transcriptase and integrase. The expression
10 cassettes described above may preserve T-helper cell and CTL epitopes. The expression cassettes may further include a polynucleotide sequence encoding an HCV *core* polypeptide, for example a sequence having at least 90% sequence identity to the sequence presented as SEQ ID
15 NO:7.

In another aspect, the invention includes an expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV *Env* polypeptide, wherein the polynucleotide sequence encoding said *Env*
20 polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59). In certain embodiments, the *Env* expression cassettes include sequences flanking a V1 region but have a deletion in the V1 region itself, for
25 example the sequence presented as SEQ ID NO:65 (Figure 52, gp160.modUS4.delV1). In certain embodiments, the *Env* expression cassettes, include sequences flanking a V2 region but have a deletion in the V2 region itself, for example the sequences shown in SEQ ID NO:60 (Figure 47);
30 SEQ ID NO:66 (Figure 53); SEQ ID NO:34 (Figure 20); SEQ ID NO:37 (Figure 24); SEQ ID NO:40 (Figure 27); SEQ ID NO:43 (Figure 30); SEQ ID NO:46 (Figure 33); SEQ ID NO:76 (Figure 64) and SEQ ID NO:49 (Figure 36). In certain

embodiments, the Env expression cassettes include sequences flanking a V1/V2 region but have a deletion in the V1/V2 region itself, for example, SEQ ID NO:59 (Figure 46); SEQ ID NO:61 (Figure 48); SEQ ID NO:67 (Figure 54); SEQ ID NO:75 (Figure 63); SEQ ID NO:35 (Figure 21); SEQ ID NO:38 (Figure 25); SEQ ID NO:41 (Figure 28); SEQ ID NO:44 (Figure 31); SEQ ID NO:47 (Figure 34) and SEQ ID NO:50 (Figure 37). The Env-encoding expression cassettes may also include a mutated cleavage site that prevents the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide, for example, SEQ ID NO:57 (Figure 44); SEQ ID NO:61 (Figure 48); SEQ ID NO:63 (Figure 50); SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34). The Env expression cassettes may include a gp160 Env polypeptide or a polypeptide derived from a gp160 Env polypeptide, for example SEQ ID NO:64 (Figure 51); SEQ ID NO:65 (Figure 52); SEQ ID NO:66 (Figure 53); SEQ ID NO:67 (Figure 54); SEQ ID NO:68 (Figure 55); SEQ ID NO:75 (Figure 63); SEQ ID NO:73 (Figure 61); SEQ ID NO:48 (Figure 35); SEQ ID NO:49 (Figure 36); SEQ ID NO:50 (Figure 37); SEQ ID NO:76 (Figure 64); and SEQ ID NO:74 (Figure 62). The Env expression cassettes may include a gp140 Env polypeptide or a polypeptide derived from a gp140 Env polypeptide, for example SEQ ID NO:56 (Figure 43); SEQ ID NO:57 (Figure 44); SEQ ID NO:58 (Figure 45); SEQ ID NO:59 (Figure 46); SEQ ID NO:60 (Figure 47); SEQ ID NO:61 (Figure 48); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); SEQ ID NO:36 (Figure 23); SEQ ID NO:37 (Figure 24); SEQ ID NO:38 (Figure 25); SEQ ID NO:39

(Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34). The Env expression cassettes may also include a gp120 Env polypeptide or a polypeptide derived from a gp120 Env polypeptide, for example SEQ ID NO:54 (Figure 41); and SEQ ID NO:55 (Figure 42); SEQ ID NO:33 (Figure 19); SEQ ID NO:34 (Figure 20); and SEQ ID NO:35 (Figure 21). The Env expression cassettes may include an Env polypeptide lacking the amino acids corresponding to residues 128 to about 194, relative to strains SF162 or US4, for example, SEQ ID NO:55 (Figure 42); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); and SEQ ID NO:68 (Figure 55).

In another aspect, the invention includes a recombinant expression system for use in a selected host cell, comprising, one or more of the expression cassettes described herein operably linked to control elements compatible with expression in the selected host cell. The expression cassettes may be included on one or on multiple vectors and may use the same or different promoters. Exemplary control elements include a transcription promoter (e.g., CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein), a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

In another aspect, the invention includes a recombinant expression system for use in a selected host cell, comprising, any one of the expression cassettes described herein operably linked to control elements

compatible with expression in the selected host cell. Exemplary control elements include, but are not limited to, a transcription promoter (e.g., CMV, CMV+intron A, SV40, RSV, HIV-LTR, MMLV-LTR, and metallothionein), a
5 transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

In yet another aspect, the invention includes a cell
10 comprising one or more of the expression cassettes described herein operably linked to control elements compatible with expression in the cell. The cell can be, for example, a mammalian cell (e.g., BHK, VERO, HT1080, 293, RD, COS-7, or CHO cells), an insect cell (e.g.,
15 *Trichoplusia ni* (Tn5) or Sf9), a bacterial cell, a plant cell, a yeast cell, an antigen presenting cell (e.g., primary, immortalized or tumor-derived lymphoid cells such as macrophages, monocytes, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof).

In another aspect, the invention includes methods
20 for producing a polypeptide including HIV *Gag*-, *prot*-, *pol*-, *reverse transcriptase*, *Env*- or *Tat*-containing polypeptide sequences, said method comprising, incubating the cells comprising one or more the expression cassettes
25 describe herein, under conditions for producing said polypeptide.

In yet another aspect, the invention includes compositions for generating an immunological response, comprising one or more of the expression cassettes
30 described herein. In certain embodiments, the compositions also include an adjuvant.

In a still further aspect, the invention includes methods of generating an immune response in a subject, comprising introducing a composition comprising one or

more of the expression cassettes described herein into the subject under conditions that are compatible with expression of said expression cassette in the subject. In certain embodiments, the expression cassette is introduced using a gene delivery vector. More than one expression cassette may be introduced using one or more gene delivery vectors.

In yet another aspect, the invention includes a purified polynucleotide comprising a polynucleotide sequence encoding a polypeptide including an HIV Env polypeptide, wherein the polynucleotide sequence encoding said Env polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59). Further exemplary purified polynucleotide sequences were presented above.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

In another embodiment, the invention includes a method for producing a polypeptide including HIV Gag polypeptide sequences, where the method comprises incubating any of the above cells containing an expression cassette of interest under conditions for producing the polypeptide.

The invention further includes, a method for producing virus-like particles (VLPs) where the method comprises incubating any of the above-described cells containing an expression cassette of interest under conditions for producing VLPs.

In another aspect the invention includes a method for producing a composition of virus-like particles (VLPs) where, any of the above-described cells containing an expression cassette of interest are incubated under

conditions for producing VLPs, and the VLPs are substantially purified to produce a composition of VLPs.

In a further embodiment of the present invention, packaging cell lines are produced using the expression cassettes of the present invention. For example, a cell line useful for packaging lentivirus vectors comprises suitable host cells that have an expression vector containing an expression cassette of the present invention wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the host cell. In a preferred embodiment, such host cells may be transfected with one or more expression cassettes having a polynucleotide sequence that encodes an HIV polymerase polypeptide or polypeptides derived therefrom, for example, where the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6. Further, the HIV polymerase polypeptide may be modified by deletions of coding regions corresponding to reverse transcriptase and integrase. Such a polynucleotide sequence may preserve T-helper cell and CTL epitopes, for example when used in a vaccine application. In addition, the polynucleotide sequence may also include other polypeptides. Further, polynucleotide sequences encoding additional polypeptides whose expression are useful for packaging cell line function may also be utilized.

In another aspect, the present invention includes a gene delivery or vaccine vector for use in a subject, where the vector is a suitable gene delivery vector for use in the subject, and the vector comprises one or more of any of the expression cassettes of the present

invention where the polynucleotide sequences of interest are operably linked to control elements compatible with expression in the subject. Such gene delivery vectors can be used in a method of DNA immunization of a subject, for example, by introducing a gene delivery vector into the subject under conditions that are compatible with expression of the expression cassette in the subject. Gene delivery vectors useful in the practice of the present invention include, but are not limited to, nonviral vectors, bacterial plasmid vectors, viral vectors, particulate carriers (where the vector is coated on a polylactide co-glycolide particles, gold or tungsten particle, for example, the coated particle can be delivered to a subject cell using a gene gun), liposome preparations, and viral vectors (e.g., vectors derived from alphaviruses, pox viruses, and vaccinia viruses, as well as, retroviral vectors, including, but not limited to, lentiviral vectors). Alphavirus-derived vectors include, for example, an alphavirus cDNA construct, a recombinant alphavirus particle preparation and a eukaryotic layered vector initiation system. In one embodiment, the subject is a vertebrate, preferably a mammal, and in a further embodiment the subject is a human.

The invention further includes a method of generating an immune response in a subject, where cells of a subject are transfected with any of the above-described gene delivery vectors (e.g., alphavirus constructs; alphavirus cDNA constructs; eukaryotic layered vector initiation systems (see, e.g., U.S. Patent Number 5,814,482 for description of suitable eukaryotic layered vector initiation systems); alphavirus particle

preparations; etc.) under conditions that permit the expression of a selected polynucleotide and production of a polypeptide of interest (i.e., encoded by any expression cassette of the present invention), thereby
5 eliciting an immunological response to the polypeptide. Transfection of the cells may be performed *ex vivo* and the transfected cells are reintroduced into the subject. Alternately, or in addition, the cells may be transfected *in vivo* in the subject. The immune response may be
10 humoral and/or cell-mediated (cellular).

Further embodiments of the present invention include purified polynucleotides. In one embodiment, the purified polynucleotide comprises a polynucleotide sequence having at least 90% sequence identity to the
15 sequence presented as SEQ ID NO:20, and complements thereof. In another embodiment, the purified polynucleotide comprises a polynucleotide sequence encoding an HIV *Gag* polypeptide, wherein the polynucleotide sequence comprises a sequence having at
20 least 90% sequence identity to the sequence presented as SEQ ID NO:20, and complements thereof. In still another embodiment, the purified polynucleotide comprises a polynucleotide sequence encoding an HIV *Gag* polypeptide, wherein the polynucleotide sequence comprises a sequence
25 having at least 90% sequence identity to the sequence presented as SEQ ID NO:9, and complements thereof. In further embodiments the polynucleotide sequence comprises a sequence having at least 90% sequence identity to one of the following sequences: SEQ ID NO:4, SEQ ID NO:5, SEQ
30 ID NO:6, SEQ ID NO:7, and complements thereof.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

These and other embodiments of the present invention will readily occur to those of ordinary skill in the art in view of the disclosure herein.

5 **BRIEF DESCRIPTION OF THE FIGURES**

Figure 1 shows the locations of the inactivation sites for the native HIV-1SF2 Gag protein coding sequence.

10 Figure 2 shows the locations of the inactivation sites for the native HIV-1SF2 Gag-protease protein coding sequence.

Figures 3A and 3B show electron micrographs of virus-like particles. Figure 3A shows immature p55Gag virus-like particles in COS-7 cells transfected with a
15 synthetic HIV-1_{SP2} *gag* construct while Figure 3B shows mature (arrows) and immature VLP in cells transfected with a modified HIV-1_{SP2} *gagprotease* construct (GP2, SEQ ID NO:70). Transfected cells were fixed at 24 h (*gag*) or 48 h (*gagprotease*) post-transfection and subsequently
20 analyzed by electron microscopy (magnification at 100,000X). Cells transfected with vector alone (pCMVMK2) served as negative control (data not shown).

Figure 4 presents an image of samples from a series of fractions which were electrophoresed on an 8-16% SDS
25 polyacrylamide gel and the resulting bands visualized by commassie blue staining. The results show that the native p55 Gag virus-like particles (VLPs) banded at a sucrose density of range of 1.15 - 1.19 g/ml with the peak at approximately 1.17 g/ml.

30 Figure 5 presents an image similar to Figure 4 where the analysis was performed using Gag VLPs produced by a synthetic Gag expression cassette.

Figure 6 presents a comparison of the total amount of purified HIV p55 Gag from several preparations obtained from two baculovirus expression cassettes encoding native and modified Gag.

5 Figure 7 presents an alignment of modified coding sequences of the present invention including a synthetic Gag expression cassette (SEQ ID NO:4), a synthetic Gag-protease expression cassette (SEQ ID NO:5), and a synthetic Gag-polymerase expression cassette (SEQ ID
10 NO:6). A common region (Gag-common; SEQ ID NO:9) extends from position 1 to position 1262.

Figure 8 presents an image of wild-type Gag-HCV core expression samples from a series of fractions which were electrophoresed on an 8-16% SDS polyacrylamide gel and
15 the resulting bands visualized by commassie staining.

Figure 9 shows the results of Western blot analysis of the gel shown presented in Figure 8.

Figure 10 presents results similar to those shown in Figure 9. The results in Figure 10 indicate that the
20 main HCV Core-specific reactivity migrates at an approximate molecular weight of 72,000 kD, which is in accordance with the predicted molecular weight of the Gag-HCV core chimeric protein.

Figures 11A to 11D present a comparison of AT
25 content, in percent, of cDNAs corresponding to an unstable human mRNA (human IFN γ mRNA; 11A), wild-type HIV Gag native RNA (11B), a stable human mRNA (human GAPDH mRNA; 11C), and synthetic HIV Gag RNA (11D).

Figure 12 shows the location of the inactivation
30 sites for the native HIV-1SF2 Gag-polymerase sequence.

Figure 13A presents a vector map of pESN2dhfr.

Figure 13B presents a map of the pCMVIII vector.

Figure 14 presents a vector map of pCMV-LINK.

Figure 15 presents a schematic diagram showing the relationships between the following forms of the HIV Env polypeptide: gp160, gp140, gp120, and gp41.

Figure 16 depicts the nucleotide sequence of wild-type gp120 from SF162 (SEQ ID NO:30).

Figure 17 depicts the nucleotide sequence of the wild-type gp140 from SF162 (SEQ ID NO:31).

Figure 18 depicts the nucleotide sequence of the wild-type gp160 from SF162 (SEQ ID NO:32).

Figure 19 depicts the nucleotide sequence of the construct designated gp120.modSF162 (SEQ ID NO:33).

Figure 20 depicts the nucleotide sequence of the construct designated gp120.modSF162.delV2 (SEQ ID NO:34).

Figure 21 depicts the nucleotide sequence of the construct designated gp120.modSF162.delV1/V2 (SEQ ID NO:35).

Figures 22A-H show the percent A-T content over the length of the sequences for IFN γ (Figures 2C and 2G); native gp160 Env US4 and SF162 (Figures 2A and 2E, respectively); GAPDH (Figures 2D and 2H); and the synthetic gp160 Env for US4 and SF162 (Figures 2B and 2F, respectively).

Figure 23 depicts the nucleotide sequence of the construct designated gp140.modSF162 (SEQ ID NO:36).

Figure 24 depicts the nucleotide sequence of the construct designated gp140.modSF162.delV2 (SEQ ID NO:37).

Figure 25 depicts the nucleotide sequence of the construct designated gp140.modSF162.delV1/V2 (SEQ ID NO:38).

Figure 26 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162 (SEQ ID NO:39).

Figure 27 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162.delV2 (SEQ ID NO:40).

Figure 28 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162.delV1/V2 (SEQ ID NO:41).

Figure 29 depicts the nucleotide sequence of the
5 construct designated gp140.mut7.modSF162 (SEQ ID NO:42).

Figure 30 depicts the nucleotide sequence of the construct designated gp140.mut7.modSF162.delV2 (SEQ ID NO:43).

Figure 31 depicts the nucleotide sequence of the
10 construct designated gp140.mut7.modSF162.delV1/V2 (SEQ ID NO:44).

Figure 32 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162 (SEQ ID NO:45).

Figure 33 depicts the nucleotide sequence of the
15 construct designated gp140.mut8.modSF162.delV2 (SEQ ID NO:46).

Figure 34 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162.delV1/V2 (SEQ ID NO:47).

Figure 35 depicts the nucleotide sequence of the
20 construct designated gp160.modSF162 (SEQ ID NO:48).

Figure 36 depicts the nucleotide sequence of the construct designated gp160.modSF162.delV2 (SEQ ID NO:49).

Figure 37 depicts the nucleotide sequence of the
25 construct designated gp160.modSF162.delV1/V2 (SEQ ID NO:50).

Figure 38 depicts the nucleotide sequence of the wild-type gp120 from US4 (SEQ ID NO:51).

Figure 39 depicts the nucleotide sequence of the
30 wild-type gp140 from US4 (SEQ ID NO:52).

Figure 40 depicts the nucleotide sequence of the wild-type gp160 from US4 (SEQ ID NO:53).

Figure 41 depicts the nucleotide sequence of the construct designated gp120.modUS4 (SEQ ID NO:54).

Figure 42 depicts the nucleotide sequence of the construct designated gpl20.modUS4.del 128-194 (SEQ ID NO:55).

5 Figure 43 depicts the nucleotide sequence of the construct designated gpl40.modUS4 (SEQ ID NO:56).

Figure 44 depicts the nucleotide sequence of the construct designated gpl40.mut.modUS4 (SEQ ID NO:57).

Figure 45 depicts the nucleotide sequence of the construct designated gpl40.TM.modUS4 (SEQ ID NO:58).

10 Figure 46 depicts the nucleotide sequence of the construct designated gpl40.modUS4.delV1/V2 (SEQ ID NO:59).

Figure 47 depicts the nucleotide sequence of the construct designated gpl40.modUS4.delV2 (SEQ ID NO:60).

15 Figure 48 depicts the nucleotide sequence of the construct designated gpl40.mut.modUS4.delV1/V2 (SEQ ID NO:61).

Figure 49 depicts the nucleotide sequence of the construct designated gpl40.modUS4.del 128-194 (SEQ ID NO:62).

20 Figure 50 depicts the nucleotide sequence of the construct designated gpl40.mut.modUS4.del 128-194 (SEQ ID NO:63).

Figure 51 depicts the nucleotide sequence of the construct designated gpl60.modUS4 (SEQ ID NO:64).

Figure 52 depicts the nucleotide sequence of the construct designated gpl60.modUS4.delV1 (SEQ ID NO:65).

Figure 53 depicts the nucleotide sequence of the construct designated gpl60.modUS4.delV2 (SEQ ID NO:66).

30 Figure 54 depicts the nucleotide sequence of the construct designated gpl60.modUS4.delV1/V2 (SEQ ID NO:67).

Figure 55 depicts the nucleotide sequence of the construct designated gpl60.modUS4.del 128-194 (SEQ ID NO:68).

5 Figure 56 depicts the nucleotide sequence of the common region of Env from wild-type US4 (SEQ ID NO:69).

Figure 57 depicts the nucleotide sequence of the common region of Env from wild-type SF162 (SEQ ID NO:70).

10 Figure 58 depicts the nucleotide sequence of synthetic sequences corresponding to the common region of Env from US4 (SEQ ID NO:71).

Figure 59 depicts the nucleotide sequence of synthetic sequences corresponding to the common region of Env from SF162 (SEQ ID NO:72).

15 Figure 60 presents a schematic representation of an Env polypeptide purification strategy.

Figure 61 depicts the nucleotide sequence of the bicistronic construct designated gpl60.modUS4.Gag.modSF2 (SEQ ID NO:73).

20 Figure 62 depicts the nucleotide sequence of the bicistronic construct designated gpl60.modSF162.Gag.modSF2 (SEQ ID NO:74).

Figure 63 depicts the nucleotide sequence of the bicistronic construct designated gpl60.modUS4.-delV1/V2.Gag.modSF2 (SEQ ID NO:75).

25 Figure 64 depicts the nucleotide sequence of the bicistronic construct designated gpl60.modSF162.delV2.Gag.modSF2 (SEQ ID NO:76).

Figures 65A-65F show micrographs of 293T cells transfected with the following polypeptide encoding sequences: Figure 65A, gag.modSF2; Figure 65B, gpl60.modUS4; Figure 65C, gpl60.modUS4.delV1/V2.gag.modSF2 (bicistronic Env and Gag); Figures 65D and 65E, gpl60.modUS4.delV1/V2 and

30

gag.modSF2; and Figure 65F, gp120.modSF162.delV2 and gag.modSF2.

Figures 66A and 66B present alignments of selected modified coding sequences of the present invention including a common region defined for each group of synthetic Env expression cassettes. Figure 66A presents alignments of modified SF162 sequences. Figure 66B presents alignments of modified US4 sequences. The SEQ ID NOs for these sequences are presented in Tables 1A and 1B.

Figure 67 shows the ELISA titers (binding antibodies) obtained in two rhesus macaques (H445, lines with solid black dots; and J408, lines with open squares). The y-axis is the end-point gp140 ELISA titers and the x-axis shows weeks post-immunization. The dashed lines at 0, 4, and 8 weeks represent DNA immunizations. The alternating dash/dotted line at 27 weeks indicates a DNA plus protein boost immunization.

Figure 68 (SEQ ID NO:77) depicts the wild-type nucleotide sequence of Gag reverse transcriptase from SF2.

Figure 69 (SEQ ID NO:78) depicts the nucleotide sequence of the construct designated GP1.

Figure 70 (SEQ ID NO:79) depicts the nucleotide sequence of the construct designated GP2.

Figure 71 (SEQ ID NO:80) depicts the nucleotide sequence of the construct designated FS(+).protinact.RTopt.YM. FS(+) indicates that there is a frameshift in the GagPol coding sequence.

Figure 72 (SEQ ID NO:81) depicts the nucleotide sequence of the construct designated FS(+).protinact.RTopt.YMWM.

Figure 73 (SEQ ID NO:82) depicts the nucleotide sequence of the construct designated FS(-)

).protmod.RTopt.YM. FS(-) indicates that there is no frameshift in the GagPol coding sequence.

Figure 74 (SEQ ID NO:83) depicts the nucleotide sequence of the construct designated

5 FS(-).protmod.RTopt.YMWM.

Figure 75 (SEQ ID NO:84) depicts the nucleotide sequence of the construct designated FS(-).protmod.RTopt(+).

10 Figure 76 (SEQ ID NO:85) depicts the nucleotide sequence of wild type Tat from isolate SF162.

Figure 77 (SEQ ID NO:86) depicts the amino acid sequence of the tat polypeptide.

Figure 78 (SEQ ID NO:87) depicts the nucleotide sequence of a synthetic Tat construct designated
15 Tat.SF162.opt.

Figure 79 (SEQ ID NO:88) depicts the nucleotide sequence of a synthetic Tat construct designated tat.cys22.sf162.opt. The construct encodes a tat polypeptide in which the cystein residue at position 22
20 of the wild type Tat polypeptide is replaced by a glycine residue.

Figures 80A to 80E are an alignment of the nucleotide sequences of the constructs designated Gag.mod.SF2, GP1 (SEQ ID NO:78), and GP2 (SEQ ID NO:79).

25 Figure 81 (SEQ ID NO:89) depicts the nucleotide sequence of the construct designated tataminoSF162.opt, which encodes the amino terminus of that tat protein. The codon encoding the cystein-22 residue is underlined.

Figure 82 (SEQ ID NO:90) depicts the amino acid
30 sequence of the polypeptide encoded by the construct designated tat.cys22.SF162.opt (SEQ ID NO:88).

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); *Methods In Enzymology* (S. Colowick and N. Kaplan, eds., Academic Press, Inc.); and *Handbook of Experimental Immunology*, Vols. I-IV (D.M. Weir and C.C. Blackwell, eds., 1986, Blackwell Scientific Publications); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Short Protocols in Molecular Biology*, 4th ed. (Ausubel et al. eds., 1999, John Wiley & Sons); *Molecular Biology Techniques: An Intensive Laboratory Course*, (Ream et al., eds., 1998, Academic Press); *PCR (Introduction to Biotechniques Series)*, 2nd ed. (Newton & Graham eds., 1997, Springer Verlag).

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more such agents.

1. DEFINITIONS

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

"Synthetic" sequences, as used herein, refers to Env-, tat- or Gag-encoding polynucleotides whose expression has been optimized as described herein, for example, by codon substitution, deletions, replacements and/or inactivation of inhibitory sequences. "Wild-type"

or "native" sequences, as used herein, refers to polypeptide encoding sequences that are essentially as they are found in nature, e.g., Gag encoding sequences as found in the isolate HIV-1SF2 or Env encoding sequences as found in the isolates HIV-1SF162 or HIV1US4.

As used herein, the term "virus-like particle" or "VLP" refers to a nonreplicating, viral shell, derived from any of several viruses discussed further below. VLPs are generally composed of one or more viral proteins, such as, but not limited to those proteins referred to as capsid, coat, shell, surface and/or envelope proteins, or particle-forming polypeptides derived from these proteins. VLPs can form spontaneously upon recombinant expression of the protein in an appropriate expression system. Methods for producing particular VLPs are known in the art and discussed more fully below. The presence of VLPs following recombinant expression of viral proteins can be detected using conventional techniques known in the art, such as by electron microscopy, biophysical characterization, and the like. See, e.g., Baker et al., *Biophys. J.* (1991) 60:1445-1456; Hagensee et al., *J. Virol.* (1994) 68:4503-4505. For example, VLPs can be isolated by density gradient centrifugation and/or identified by characteristic density banding (e.g., Example 7). Alternatively, cryoelectron microscopy can be performed on vitrified aqueous samples of the VLP preparation in question, and images recorded under appropriate exposure conditions.

By "particle-forming polypeptide" derived from a particular viral protein is meant a full-length or near full-length viral protein, as well as a fragment thereof, or a viral protein with internal deletions, which has the ability to form VLPs under conditions that favor VLP

- formation. Accordingly, the polypeptide may comprise the full-length sequence, fragments, truncated and partial sequences, as well as analogs and precursor forms of the reference molecule. The term therefore intends
- 5 deletions, additions and substitutions to the sequence, so long as the polypeptide retains the ability to form a VLP. Thus, the term includes natural variations of the specified polypeptide since variations in coat proteins often occur between viral isolates. The term also
- 10 includes deletions, additions and substitutions that do not naturally occur in the reference protein, so long as the protein retains the ability to form a VLP. Preferred substitutions are those which are conservative in nature, i.e., those substitutions that take place within a family
- 15 of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic -- aspartate and glutamate; (2) basic -- lysine, arginine, histidine; (3) non-polar -- alanine, valine, leucine, isoleucine, proline,
- 20 phenylalanine, methionine, tryptophan; and (4) uncharged polar -- glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids.
- 25 An "antigen" refers to a molecule containing one or more epitopes (either linear, conformational or both) that will stimulate a host's immune system to make a humoral and/or cellular antigen-specific response. The term is used interchangeably with the term "immunogen."
- 30 Normally, a B-cell epitope will include at least about 5 amino acids but can be as small as 3-4 amino acids. A T-cell epitope, such as a CTL epitope, will include at least about 7-9 amino acids, and a helper T-cell epitope at least about 12-20 amino acids. Normally, an epitope

will include between about 7 and 15 amino acids, such as, 9, 10, 12 or 15 amino acids. The term "antigen" denotes both subunit antigens, (i.e., antigens which are separate and discrete from a whole organism with which the antigen is associated in nature), as well as, killed, attenuated or inactivated bacteria, viruses, fungi, parasites or other microbes. Antibodies such as anti-idiotypic antibodies, or fragments thereof, and synthetic peptide mimotopes, which can mimic an antigen or antigenic determinant, are also captured under the definition of antigen as used herein. Similarly, an oligonucleotide or polynucleotide which expresses an antigen or antigenic determinant *in vivo*, such as in gene therapy and DNA immunization applications, is also included in the definition of antigen herein.

For purposes of the present invention, antigens can be derived from any of several known viruses, bacteria, parasites and fungi, as described more fully below. The term also intends any of the various tumor antigens. Furthermore, for purposes of the present invention, an "antigen" refers to a protein which includes modifications, such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the protein maintains the ability to elicit an immunological response, as defined herein. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the antigens.

An "immunological response" to an antigen or composition is the development in a subject of a humoral and/or a cellular immune response to an antigen present in the composition of interest. For purposes of the present invention, a "humoral immune response" refers to

an immune response mediated by antibody molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important aspect of cellular immunity involves an antigen-specific response by cytolytic T-cells ("CTL"s). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the surfaces of cells. CTLs help induce and promote the destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood cells, including those derived from CD4+ and CD8+ T-cells.

A composition or vaccine that elicits a cellular immune response may serve to sensitize a vertebrate subject by the presentation of antigen in association with MHC molecules at the cell surface. The cell-mediated immune response is directed at, or near, cells presenting antigen at their surface. In addition, antigen-specific T-lymphocytes can be generated to allow for the future protection of an immunized host.

The ability of a particular antigen to stimulate a cell-mediated immunological response may be determined by a number of assays, such as by lymphoproliferation (lymphocyte activation) assays, CTL cytotoxic cell assays, or by assaying for T-lymphocytes specific for the

antigen in a sensitized subject. Such assays are well known in the art. See, e.g., Erickson et al., *J. Immunol.* (1993) 151:4189-4199; Doe et al., *Eur. J. Immunol.* (1994) 24:2369-2376. Recent methods of measuring cell-mediated immune response include measurement of intracellular cytokines or cytokine secretion by T-cell populations, or by measurement of epitope specific T-cells (e.g., by the tetramer technique) (reviewed by McMichael, A.J., and O'Callaghan, C.A., *J. Exp. Med.* 187(9)1367-1371, 1998; Mcheyzer-Williams, M.G., et al, *Immunol. Rev.* 150:5-21, 1996; Lalvani, A., et al, *J. Exp. Med.* 186:859-865, 1997).

Thus, an immunological response as used herein may be one which stimulates the production of CTLs, and/or the production or activation of helper T- cells. The antigen of interest may also elicit an antibody-mediated immune response. Hence, an immunological response may include one or more of the following effects: the production of antibodies by B-cells; and/or the activation of suppressor T-cells and/or $\gamma\delta$ T-cells directed specifically to an antigen or antigens present in the composition or vaccine of interest. These responses may serve to neutralize infectivity, and/or mediate antibody-complement, or antibody dependent cell cytotoxicity (ADCC) to provide protection to an immunized host. Such responses can be determined using standard immunoassays and neutralization assays, well known in the art.

An "immunogenic composition" is a composition that comprises an antigenic molecule where administration of the composition to a subject results in the development in the subject of a humoral and/or a cellular immune response to the antigenic molecule of interest.

By "subunit vaccine" is meant a vaccine composition which includes one or more selected antigens but not all antigens, derived from or homologous to, an antigen from a pathogen of interest such as from a virus, bacterium, 5 parasite or fungus. Such a composition is substantially free of intact pathogen cells or pathogenic particles, or the lysate of such cells or particles. Thus, a "subunit vaccine" can be prepared from at least partially purified (preferably substantially purified) immunogenic 10 polypeptides from the pathogen, or analogs thereof. The method of obtaining an antigen included in the subunit vaccine can thus include standard purification techniques, recombinant production, or synthetic production.

15 "Substantially purified" general refers to isolation of a substance (compound, polynucleotide, protein, polypeptide, polypeptide composition) such that the substance comprises the majority percent of the sample in which it resides. Typically in a sample a substantially 20 purified component comprises 50%, preferably 80%-85%, more preferably 90-95% of the sample. Techniques for purifying polynucleotides and polypeptides of interest are well-known in the art and include, for example, ion-exchange chromatography, affinity chromatography and 25 sedimentation according to density.

A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* when placed 30 under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but

is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral or procaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence may be located 3' to the coding sequence.

Typical "control elements", include, but are not limited to, transcription promoters, transcription enhancer elements, transcription termination signals, polyadenylation sequences (located 3' to the translation stop codon), sequences for optimization of initiation of translation (located 5' to the coding sequence), and translation termination sequences, see e.g., McCaughan et al. (1995) *PNAS USA* 92:5431-5435; Kochetov et al (1998) *FEBS Letts.* 440:351-355.

A "nucleic acid" molecule can include, but is not limited to, procaryotic sequences, eucaryotic mRNA, cDNA from eucaryotic mRNA, genomic DNA sequences from eucaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. The term also captures sequences that include any of the known base analogs of DNA and RNA.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a given promoter operably linked to a coding sequence is capable of effecting the expression of the coding sequence when the proper enzymes are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. "Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such terms denoting procaryotic microorganisms or eucaryotic cell lines cultured as unicellular entities, are used interchangeably, and refer to cells which can be, or have been, used as recipients for recombinant vectors or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms.

Techniques for determining amino acid sequence "similarity" are well known in the art. In general, "similarity" means the exact amino acid to amino acid comparison of two or more polypeptides at the appropriate place, where amino acids are identical or possess similar chemical and/or physical properties such as charge or hydrophobicity. A so-termed "percent similarity" then

can be determined between the compared polypeptide sequences. Techniques for determining nucleic acid and amino acid sequence identity also are well known in the art and include determining the nucleotide sequence of the mRNA for that gene (usually via a cDNA intermediate) and determining the amino acid sequence encoded thereby, and comparing this to a second amino acid sequence. In general, "identity" refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of two polynucleotides or polypeptide sequences, respectively.

Two or more polynucleotide sequences can be compared by determining their "percent identity." Two or more amino acid sequences likewise can be compared by determining their "percent identity." The percent identity of two sequences, whether nucleic acid or peptide sequences, is generally described as the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100.

An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981). This algorithm can be extended to use with peptide sequences using the scoring matrix developed by Dayhoff, *Atlas of Protein Sequences and Structure*, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). An implementation of this algorithm for nucleic acid and peptide sequences is provided by the Genetics Computer Group (Madison, WI) in their BestFit utility application. The default parameters for this method are

described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). Other equally suitable programs for calculating the percent identity or
5 similarity between sequences are generally known in the art.

For example, percent identity of a particular nucleotide sequence to a reference sequence can be determined using the homology algorithm of Smith and
10 Waterman with a default scoring table and a gap penalty of six nucleotide positions. Another method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh,
15 developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages, the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap
20 extension penalty of one, and a gap of six). From the data generated, the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, such as the alignment program
25 BLAST, which can also be used with default parameters. For example, BLASTN and BLASTP can be used with the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by =
30 HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at

the following internet address:

<http://www.ncbi.nlm.gov/cgi-bin/BLAST>.

One of skill in the art can readily determine the proper search parameters to use for a given sequence in the above programs. For example, the search parameters may vary based on the size of the sequence in question. Thus, for example, a representative embodiment of the present invention would include an isolated polynucleotide having X contiguous nucleotides, wherein

10 (i) the X contiguous nucleotides have at least about 50% identity to Y contiguous nucleotides derived from any of the sequences described herein, (ii) X equals Y, and (iii) X is greater than or equal to 6 nucleotides and up to 5000 nucleotides, preferably greater than or equal to

15 8 nucleotides and up to 5000 nucleotides, more preferably 10-12 nucleotides and up to 5000 nucleotides, and even more preferably 15-20 nucleotides, up to the number of nucleotides present in the full-length sequences described herein (e.g., see the Sequence Listing and

20 claims), including all integer values falling within the above-described ranges.

The synthetic expression cassettes (and purified polynucleotides) of the present invention include related polynucleotide sequences having about 80% to 100%,

25 greater than 80-85%, preferably greater than 90-92%, more preferably greater than 95%, and most preferably greater than 98% sequence (including all integer values falling within these described ranges) identity to the synthetic expression cassette sequences disclosed herein (for

30 example, to the sequences presented in Tables 1A and 1B) when the sequences of the present invention are used as the query sequence.

Two nucleic acid fragments are considered to "selectively hybridize" as described herein. The degree of sequence identity between two nucleic acid molecules affects the efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit a completely identical sequence from hybridizing to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A nucleic acid molecule that is capable of hybridizing selectively to a target sequence under "moderately stringent" typically

hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected
5 nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe.
10 Hybridization conditions useful for probe/target hybridization where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames
15 and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a
20 particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or
25 absence of blocking agents in the hybridization solutions (e.g., formamide, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions
30 is selected following standard methods in the art (see, for example, Sambrook, et al., Molecular Cloning: A

Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.).

5 A first polynucleotide is "derived from" second polynucleotide if it has the same or substantially the same basepair sequence as a region of the second polynucleotide, its cDNA, complements thereof, or if it displays sequence identity as described above.

10 A first polypeptide is "derived from" a second polypeptide if it is (i) encoded by a first polynucleotide derived from a second polynucleotide, or (ii) displays sequence identity to the second polypeptides as described above.

15 Generally, a viral polypeptide is "derived from" a particular polypeptide of a virus (viral polypeptide) if it is (i) encoded by an open reading frame of a polynucleotide of that virus (viral polynucleotide), or (ii) displays sequence identity to polypeptides of that virus as described above.

20 "Encoded by" refers to a nucleic acid sequence which codes for a polypeptide sequence, wherein the polypeptide sequence or a portion thereof contains an amino acid sequence of at least 3 to 5 amino acids, more preferably at least 8 to 10 amino acids, and even more preferably at least 15 to 20 amino acids from a polypeptide encoded by
25 the nucleic acid sequence. Also encompassed are polypeptide sequences which are immunologically identifiable with a polypeptide encoded by the sequence.

"Purified polynucleotide" refers to a polynucleotide of interest or fragment thereof which is essentially
30 free, e.g., contains less than about 50%, preferably less than about 70%, and more preferably less than about 90%, of the protein with which the polynucleotide is naturally associated. Techniques for purifying polynucleotides of interest are well-known in the art and include, for

example, disruption of the cell containing the polynucleotide with a chaotropic agent and separation of the polynucleotide(s) and proteins by ion-exchange chromatography, affinity chromatography and sedimentation
5 according to density.

By "nucleic acid immunization" is meant the introduction of a nucleic acid molecule encoding one or more selected antigens into a host cell, for the *in vivo* expression of an antigen, antigens, an epitope, or
10 epitopes. The nucleic acid molecule can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal and mucosal administration, or the like, or can be introduced *ex vivo*, into cells which have been removed from the host. In the latter
15 case, the transformed cells are reintroduced into the subject where an immune response can be mounted against the antigen encoded by the nucleic acid molecule.

"Gene transfer" or "gene delivery" refers to methods or systems for reliably inserting DNA or RNA of interest
20 into a host cell. Such methods can result in transient expression of non-integrated transferred DNA, extrachromosomal replication and expression of transferred replicons (e.g., episomes), or integration of transferred genetic material into the genomic DNA of host
25 cells. Gene delivery expression vectors include, but are not limited to, vectors derived from bacterial plasmid vectors, viral vectors, non-viral vectors, alphaviruses, pox viruses and vaccinia viruses. When used for immunization, such gene delivery expression vectors may
30 be referred to as vaccines or vaccine vectors.

"T lymphocytes" or "T cells" are non-antibody producing lymphocytes that constitute a part of the cell-mediated arm of the immune system. T cells arise from immature lymphocytes that migrate from the bone marrow to

the thymus, where they undergo a maturation process under the direction of thymic hormones. Here, the mature lymphocytes rapidly divide increasing to very large numbers. The maturing T cells become immunocompetent
5 based on their ability to recognize and bind a specific antigen. Activation of immunocompetent T cells is triggered when an antigen binds to the lymphocyte's surface receptors.

The term "transfection" is used to refer to the
10 uptake of foreign DNA by a cell. A cell has been "transfected" when exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are generally known in the art. See, e.g., Graham et al. (1973) *Virology*, 52:456, Sambrook et al.
15 (1989) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratories, New York, Davis et al. (1986) *Basic Methods in Molecular Biology*, Elsevier, and Chu et al. (1981) *Gene* 13:197. Such techniques can be used to introduce one or more exogenous DNA moieties into
20 suitable host cells. The term refers to both stable and transient uptake of the genetic material, and includes uptake of peptide- or antibody-linked DNAs.

A "vector" is capable of transferring gene sequences to target cells (e.g., bacterial plasmid vectors, viral
25 vectors, non-viral vectors, particulate carriers, and liposomes). Typically, "vector construct," "expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to
30 target cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

Transfer of a "suicide gene" (e.g., a drug-susceptibility gene) to a target cell renders the cell sensitive to compounds or compositions that are

relatively nontoxic to normal cells. Moolten, F.L. (1994) *Cancer Gene Ther.* 1:279-287. Examples of suicide genes are thymidine kinase of herpes simplex virus (HSV-tk), cytochrome P450 (Manome et al. (1996) *Gene Therapy* 3:513-520), human deoxycytidine kinase (Manome et al. (1996) *Nature Medicine* 2(5):567-573) and the bacterial enzyme cytosine deaminase (Dong et al. (1996) *Human Gene Therapy* 7:713-720). Cells which express these genes are rendered sensitive to the effects of the relatively nontoxic prodrugs ganciclovir (HSV-tk), cyclophosphamide (cytochrome P450 2B1), cytosine arabinoside (human deoxycytidine kinase) or 5-fluorocytosine (bacterial cytosine deaminase). Culver et al. (1992) *Science* 256:1550-1552, Huber et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:8302-8306.

A "selectable marker" or "reporter marker" refers to a nucleotide sequence included in a gene transfer vector that has no therapeutic activity, but rather is included to allow for simpler preparation, manufacturing, characterization or testing of the gene transfer vector.

A "specific binding agent" refers to a member of a specific binding pair of molecules wherein one of the molecules specifically binds to the second molecule through chemical and/or physical means. One example of a specific binding agent is an antibody directed against a selected antigen.

By "subject" is meant any member of the subphylum chordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such

as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered. The system described above
5 is intended for use in any of the above vertebrate species, since the immune systems of all of these vertebrates operate similarly.

By "pharmaceutically acceptable" or "pharmacologically acceptable" is meant a material which
10 is not biologically or otherwise undesirable, i.e., the material may be administered to an individual in a formulation or composition without causing any undesirable biological effects or interacting in a deleterious manner with any of the components of the
15 composition in which it is contained.

By "physiological pH" or a "pH in the physiological range" is meant a pH in the range of approximately 7.2 to 8.0 inclusive, more typically in the range of approximately 7.2 to 7.6 inclusive.

20 As used herein, "treatment" refers to any of (i) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the pathogen in question. Treatment may
25 be effected prophylactically (prior to infection) or therapeutically (following infection).

"Lentiviral vector", and "recombinant lentiviral vector" are derived from the subset of retroviral vectors known as lentiviruses. Lentiviral vectors refer to a
30 nucleic acid construct which carries, and within certain embodiments, is capable of directing the expression of a nucleic acid molecule of interest. The lentiviral vector includes at least one transcriptional promoter/enhancer or locus defining element(s), or other elements which

control gene expression by other means such as alternate splicing, nuclear RNA export, post-translational modification of messenger, or post-transcriptional modification of protein. Such vector constructs must
5 also include a packaging signal, long terminal repeats (LTRS) or portion thereof, and positive and negative strand primer binding sites appropriate to the lentiviral vector used (if these are not already present in the retroviral vector). Optionally, the recombinant
10 lentiviral vector may also include a signal which directs polyadenylation, selectable markers such as Neo, TK, hygromycin, phleomycin, histidinol, or DHFR, as well as one or more restriction sites and a translation termination sequence. By way of example, such vectors
15 typically include a 5' LTR, a tRNA binding site, a packaging signal, an origin of second strand DNA synthesis, and a 3'LTR or a portion thereof.

"Lentiviral vector particle" as utilized within the present invention refers to a lentivirus which carries at
20 least one gene of interest. The retrovirus may also contain a selectable marker. The recombinant lentivirus is capable of reverse transcribing its genetic material (RNA) into DNA and incorporating this genetic material into a host cell's DNA upon infection. Lentiviral vector
25 particles may have a lentiviral envelope, a non-lentiviral envelope (e.g., an amphi or VSV-G envelope), or a chimeric envelope.

"Nucleic acid expression vector" or "Expression cassette" refers to an assembly which is capable of
30 directing the expression of a sequence or gene of interest. The nucleic acid expression vector includes a promoter which is operably linked to the sequences or gene(s) of interest. Other control elements may be present as well. Expression cassettes described herein

may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also include a bacterial origin of replication, one or more selectable markers, a signal which allows the plasmid construct to exist as single-stranded DNA (e.g., a M13 origin of replication), a multiple cloning site, and a "mammalian" origin of replication (e.g., a SV40 or adenovirus origin of replication).

"Packaging cell" refers to a cell which contains those elements necessary for production of infectious recombinant retrovirus (e.g., lentivirus) which are lacking in a recombinant retroviral vector. Typically, such packaging cells contain one or more expression cassettes which are capable of expressing proteins which encode *Gag*, *pol* and *env* proteins.

"Producer cell" or "vector producing cell" refers to a cell which contains all elements necessary for production of recombinant retroviral vector particles.

2. MODES OF CARRYING OUT THE INVENTION

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular formulations or process parameters as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

Although a number of methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

2.1 SYNTHETIC EXPRESSION CASSETTES

2.1.1 MODIFICATION OF HIV-1 GAG NUCLEIC ACID CODING SEQUENCES

One aspect of the present invention is the generation of HIV-1 Gag protein coding sequences, and related sequences, having improved expression relative to the corresponding wild-type sequence. An exemplary embodiment of the present invention is illustrated herein modifying the Gag protein wild-type sequences obtained from the HIV-1SF2 strain (SEQ ID NO:1; Sanchez-Pescador, R., et al., *Science* 227(4686): 484-492, 1985; Luciw, P.A., et al. U.S. Patent No. 5,156,949, issued October 20, 1992; Luciw, P.A., et al., U.S. Patent No. 5,688,688, November 18, 1997). Gag sequence obtained from other HIV variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Gag protein encoding sequences obtained from the isolates HIV_{IIIb}, HIV_{SF2}, HIV-1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LAI}, HIV_{NIH}, HIV-1_{CM235}, HIV-1_{US4}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2_{UC1} and HIV-2_{UC2}), and simian immunodeficiency virus (SIV). (See, e.g., *Virology*, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991); *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA; for a description of these and other related viruses).

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes (Example 1). The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet.

The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Gag coding sequences were modified to be comparable to codon usage found in highly expressed human genes. In Figure 11 (Example 1), the percent A-T content of cDNA sequences corresponding to the mRNA for a known unstable mRNA and a known stable mRNA are compared to the percent A-T content of native HIV-1SF2 Gag cDNA and to the synthetic Gag cDNA sequence of the present invention. Experiments performed in support of the present invention showed that the synthetic Gag sequences were capable of higher level of protein production (see the Examples) relative to the native Gag sequences. The data in Figure 11 suggest that one reason for this increased production is increased stability of the mRNA corresponding to the synthetic Gag coding sequences versus the mRNA corresponding to the native Gag coding sequences.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag coding sequences (Example 1). The RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements were inactivated by introducing multiple point mutations that did not alter the reading frame of the encoded proteins. Figure 1 shows the original SF2 Gag sequence, the location of the INS sequences, and the modifications made to the INS sequences to reduce their effects. The resulting modified coding sequences are

presented as a synthetic Gag expression cassette (SEQ ID NO:4).

Modification of the Gag polypeptide coding sequences resulted in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Further, expression of the sequences resulted in production of virus-like particles (VLPs) by these cell lines (see below). Similar Gag polypeptide coding sequences can be obtained from a variety of isolates (families, sub-types, strains, etc.) including, but not limited to such other variants include, but are not limited to, Gag polypeptide encoding sequences obtained from the isolates HIV_{IIIB}, HIV₈₇₂, HIV-1_{SP162}, HIV-1_{SP170}, HIV_{LAV}, HIV_{LAI}, HIV₉₀, HIV-1_{CH235}, HIV-1_{US4}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2_{UC1} and HIV-2_{UC2}), and simian immunodeficiency virus (SIV). (See, e.g., Virology, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991; *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA). Gag polypeptide encoding sequences derived from these variants can be optimized and tested for improved expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 1).

2.1.2 FURTHER MODIFICATION OF SEQUENCES INCLUDING HIV-1 GAG NUCLEIC ACID CODING SEQUENCES

Experiments performed in support of the present invention have shown that similar modifications of HIV-1 Gag-protease, Gag-reverse transcriptase and Gag-polymerase sequences also result in improved expression

of the polyproteins, as well as, the production of VLPs formed by polypeptides produced from such modified coding sequences.

For the Gag-protease sequence (wild type, SEQ ID NO:2; modified, SEQ ID NOs:5, 78, 79), the changes in
5 codon usage were restricted to the regions upstream of the -1 frameshift (Figure 2). Further, inhibitory (or instability) elements (INS) located within the coding sequences of the Gag-protease polypeptide coding sequence
10 were altered as well (indicated in Figure 2). Exemplary constructs (which include the -1 frameshift) encoding modified Gag-protease sequences include those shown in SEQ ID NOs:78 and 79 (Figures 69 and 70). These are: GP1 (SEQ ID NO:78) in which the protease region was also
15 codon optimized and INS inactivated and GP2 (SEQ ID NO:79), in which the protease region was only subjected to INS inactivation.

For other Gag-containing sequences, for example the Gag-polymerase sequence (wild type, SEQ ID NO:3; modified, SEQ ID NO:6) or Gag-reverse transcriptase (wild
20 type, SEQ ID NO:77; modified SEQ ID NOs:80-84), the changes in codon usage are similar to those for the Gag-protease sequence. Those expression cassettes which contain a frameshift in the GagPol coding sequence are
25 designated "FS(+)" (SEQ ID NOs:80 and 81, Figures 71 and 72) while the designation "FS(-)" (SEQ ID NOs: 82, 83 and 84, Figures 73, 74 and 75) indicates that there is no frameshift utilized in this coding sequence.

In addition to polyproteins containing HIV-related
30 sequences, the various Gag-, Gag-prot, Gag-pol, Gag-reverse transcriptase encoding sequences of the present invention can be fused to other polypeptides (creating chimeric polypeptides) for which an immunogenic response is desired. An example of such a chimeric protein is the

joining of the improved expression Gag encoding sequences to the Hepatitis C Virus (HCV) core protein. In this case, the HCV-core encoding sequences were placed in-frame with the HIV-Gag encoding sequences, resulting in the Gag/HCV-core encoding sequence presented as SEQ ID NO:7 (wild type sequence presented as SEQ ID NO:8).

Further sequences useful in the practice of the present invention include, but are not limited to, sequences encoding viral epitopes/antigens (including but not limited to, HCV antigens (e.g., E1, E2; Houghton, M., et al., U.S. Patent No. 5,714,596, issued February 3, 1998; Houghton, M., et al., U.S. Patent No. 5,712,088, issued January 27, 1998; Houghton, M., et al., U.S. Patent No. 5,683,864, issued November 4, 1997; Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J., et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No. 5,670,152, issued September 23, 1997), HIV antigens (e.g., derived from *nef*, *tat*, *rev*, *vpu*, *vif*, *vpr* and/or *env*); and sequences encoding tumor antigens/epitopes. Additional sequences are described below. Also, variations on the orientation of the Gag and other coding sequences, relative to each other, are also described below.

Gag, Gag-protease, Gag-reverse transcriptase and/or Gag-polymerase polypeptide coding sequences can be obtained from any HIV isolates (different families, subtypes, and strains) including but not limited to the isolates HIV_{111b}, HIV_{SP2}, HIV_{SP162}, HIV_{us4}, HIV_{cm235}, HIV_{LAV}, HIV_{LAT}, HIV₉₈ (see, e.g., Myers et al. Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos, New Mexico: Los Alamos National Laboratory). Synthetic expression cassettes can be generated using

such coding sequences as starting material by following the teachings of the present specification (e.g., see Example 1). Further, the synthetic expression cassettes of the present invention include related Gag polypeptide coding sequences having greater than 75%, preferably greater than 80-85%, more preferably greater than 90-95%, and most preferably greater than 98% sequence identity (or any integer value within these ranges) to the synthetic expression cassette sequences disclosed herein (for example, SEQ ID NO:4; SEQ ID NO:5; SEQ ID NO:6; and SEQ ID NO:20, the Gag Major Homology Region).

2.1.3 EXPRESSION OF SYNTHETIC SEQUENCES ENCODING HIV-1 GAG AND RELATED POLYPEPTIDES

Several synthetic Gag-encoding sequences (expression cassettes) of the present invention were cloned into a number of different expression vectors (Example 1) to evaluate levels of expression and production of VLPs. Two modified synthetic coding sequences are presented as a synthetic Gag expression cassette (SEQ ID NO:4) and a synthetic Gag-protease expression cassette (SEQ ID NOs:78 and 79). Other synthetic Gag-encoding proteins are presented, for example, as SEQ ID NOs:80 through 84. The synthetic DNA fragments for Gag-encoding polypeptides (e.g., Gag, Gag-protease, Gag-polymerase, Gag-reverse transcriptase) were cloned into expression vectors described in Example 1, including, a transient expression vector, CMV-promoter-based mammalian vectors, and a shuttle vector for use in baculovirus expression systems. Corresponding wild-type sequences were cloned into the same vectors.

These vectors were then transfected into a several different cell types, including a variety of mammalian

cell lines, (293, RD, COS-7, and CHO, cell lines available, for example, from the A.T.C.C.). The cell lines were cultured under appropriate conditions and the levels of p24 (Gag) expression in supernatants were
5 evaluated (Example 2). The results of these assays demonstrated that expression of synthetic Gag-encoding sequences were significantly higher than corresponding wild-type sequences (Example 2; Table 2).

Further, Western Blot analysis showed that cells
10 containing the synthetic Gag expression cassette produced the expected 55 kD (p55) protein at higher per-cell concentrations than cells containing the native expression cassette. The Gag p55 protein was seen in both cell lysates and supernatants. The levels of
15 production were significantly higher in cell supernatants for cells transfected with the synthetic Gag expression cassette of the present invention. Experiments performed in support of the present invention suggest that cells containing the synthetic Gag-prot expression cassettes
20 produced the expected Gag-prot protein at comparably higher per-cell concentrations than cells containing the wild-type expression cassette.

Fractionation of the supernatants from mammalian cells transfected with the synthetic Gag expression
25 cassette showed that it provides superior production of both p55 protein and VLPs, relative to the wild-type Gag sequences (Examples 6 and 7).

Efficient expression of these Gag-containing polypeptides in mammalian cell lines provides the
30 following benefits: the Gag polypeptides are free of baculovirus contaminants; production by established methods approved by the FDA; increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Gag-containing polypeptides in

CHO or other mammalian cells which is not feasible in the absence of the increased expression obtained using the constructs of the present invention. Exemplary Mammalian cell lines include, but are not limited to, BHK, VERO, HT1080, 293, 293T, RD, COS-7, CHO, Jurkat, HUT, SUPT, C8166, MOLT4/clone8, MT-2, MT-4, H9, PM1, CEM, myeloma cells (e.g., SB20 cells) and CEMX174, such cell lines are available, for example, from the A.T.C.C.).

A synthetic Gag expression cassette of the present invention also demonstrated high levels of expression and VLP production when transfected into insect cells (Example 7). Further, in addition to a higher total protein yield, the final product from the synthetic p55-expressed Gag consistently contained lower amounts of contaminating baculovirus proteins than the final purified product from the native p55-expressed Gag.

Further, synthetic Gag expression cassettes of the present invention have also been introduced into yeast vectors which were transformed into and efficiently expressed by yeast cells (*Saccharomyces cerevisiae*; using vectors as described in Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998).

In addition to the mammalian and insect vectors described in the Examples, the synthetic expression cassettes of the present invention can be incorporated into a variety of expression vectors using selected expression control elements. Appropriate vectors and control elements for any given cell type can be selected by one having ordinary skill in the art in view of the teachings of the present specification and information known in the art about expression vectors.

For example, a synthetic Gag expression cassette can be inserted into a vector which includes control elements operably linked to the desired coding sequence, which

allow for the expression of the gene in a selected cell-type. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter (a CMV promoter can include intron A), RSV, HIV-LTR, the mouse mammary tumor virus LTR promoter (MMLV-LTR), FIV-LTR, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook, et al., *supra*, as well as a bovine growth hormone terminator sequence. Introns, containing splice donor and acceptor sites, may also be designed into the constructs for use with the present invention (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

- The desired synthetic Gag polypeptide encoding sequences can be cloned into any number of commercially available vectors to generate expression of the polypeptide in an appropriate host system. These systems include, but are not limited to, the following:
- 5 baculovirus expression {Reilly, P.R., et al., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992); Beames, et al., *Biotechniques* 11:378 (1991); Pharmingen; Clontech, Palo Alto, CA}}, vaccinia expression {Earl, P. L., et al.,
- 10 "Expression of proteins in mammalian cells using vaccinia" In *Current Protocols in Molecular Biology* (F. M. Ausubel, et al. Eds.), Greene Publishing Associates & Wiley Interscience, New York (1991); Moss, B., et al., U.S. Patent Number 5,135,855, issued 4 August 1992},
- 15 expression in bacteria {Ausubel, F.M., et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley and Sons, Inc., Media PA; Clontech}, expression in yeast {Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998; Shuster, J.R., U.S. Patent No. 5,629,203,
- 20 issued May 13, 1997; Gellissen, G., et al., *Antonie Van Leeuwenhoek*, 62(1-2):79-93 (1992); Romanos, M.A., et al., *Yeast* 8(6):423-488 (1992); Goeddel, D.V., *Methods in Enzymology* 185 (1990); Guthrie, C., and G.R. Fink, *Methods in Enzymology* 194 (1991)}, expression in
- 25 mammalian cells {Clontech; Gibco-BRL, Ground Island, NY; e.g., Chinese hamster ovary (CHO) cell lines (Haynes, J., et al., *Nuc. Acid. Res.* 11:687-706 (1983); 1983, Lau, Y.F., et al., *Mol. Cell. Biol.* 4:1469-1475 (1984); Kaufman, R. J., "Selection and coamplification of
- 30 heterologous genes in mammalian cells," in *Methods in Enzymology*, vol. 185, pp537-566. Academic Press, Inc., San Diego CA (1991)}, and expression in plant cells {plant cloning vectors, Clontech Laboratories, Inc., Palo Alto, CA, and Pharmacia LKB Biotechnology, Inc.,

- Piscataway, NJ; Hood, E., et al., *J. Bacteriol.* 168:1291-1301 (1986); Nagel, R., et al., *FEMS Microbiol. Lett.* 67:325 (1990); An, et al., "Binary Vectors", and others in Plant Molecular Biology Manual A3:1-19 (1988);
- 5 Miki, B.L.A., et al., pp.249-265, and others in Plant DNA Infectious Agents (Hohn, T., et al., eds.) Springer-Verlag, Wien, Austria, (1987); *Plant Molecular Biology: Essential Techniques*, P.G. Jones and J.M. Sutton, New York, J. Wiley, 1997; Miglani, Gurbachan *Dictionary of*
- 10 *Plant Genetics and Molecular Biology*, New York, Food Products Press, 1998; Henry, R. J., *Practical Applications of Plant Molecular Biology*, New York, Chapman & Hall, 1997}.

- Also included in the invention is an expression
- 15 vector, such as the CMV promoter-containing vectors described in Example 1, containing coding sequences and expression control elements which allow expression of the coding regions in a suitable host. The control elements generally include a promoter, translation initiation
- 20 codon, and translation and transcription termination sequences, and an insertion site for introducing the insert into the vector. Translational control elements have been reviewed by M. Kozak (e.g., Kozak, M., *Mamm. Genome* 7(8):563-574, 1996; Kozak, M., *Biochimie*
- 25 76(9):815-821, 1994; Kozak, M., *J Cell Biol* 108(2):229-241, 1989; Kozak, M., and Shatkin, A.J., *Methods Enzymol* 60:360-375, 1979).

- Expression in yeast systems has the advantage of commercial production. Recombinant protein production by
- 30 vaccinia and CHO cell line have the advantage of being mammalian expression systems. Further, vaccinia virus expression has several advantages including the following: (i) its wide host range; (ii) faithful post-

transcriptional modification, processing, folding, transport, secretion, and assembly of recombinant proteins; (iii) high level expression of relatively soluble recombinant proteins; and (iv) a large capacity to accommodate foreign DNA.

The recombinantly expressed polypeptides from synthetic Gag-encoding expression cassettes are typically isolated from lysed cells or culture media. Purification can be carried out by methods known in the art including salt fractionation, ion exchange chromatography, gel filtration, size-exclusion chromatography, size-fractionation, and affinity chromatography. Immunoaffinity chromatography can be employed using antibodies generated based on, for example, Gag antigens.

Advantages of expressing the Gag-containing proteins of the present invention using mammalian cells include, but are not limited to, the following: well-established protocols for scale-up production; the ability to produce VLPs; cell lines are suitable to meet good manufacturing process (GMP) standards; culture conditions for mammalian cells are known in the art.

2.1.4 MODIFICATION OF HIV-1 ENV NUCLEIC ACID CODING SEQUENCES

One aspect of the present invention is the generation of HIV-1 Env protein coding sequences, and related sequences, having improved expression relative to the corresponding wild-type sequence. Exemplary embodiments of the present invention are illustrated herein modifying the Env protein wild-type sequences obtained from the HIV-1 subtype B strains HIV-1US4 and HIV-1SF162 (Myers et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos,

New Mexico: Los Alamos National Laboratory). Env sequence obtained from other HIV variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include those
5 described above in Section 2.1.1 and on the World Wide Web (Internet), for example at <http://hiv-web.lanl.gov/cgi-bin/hivDB3/public/wdb/ssampublic> and <http://hiv-web.lanl.gov>.

First, the HIV-1 codon usage pattern was modified so
10 that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes (Example 1). The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content
15 in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Env coding sequences were modified to be comparable to codon usage found in highly
20 expressed human genes. Experiments performed in support of the present invention showed that the synthetic Env sequences were capable of higher level of protein production (see the Examples) relative to the native Env sequences. One reason for this increased production may
25 be increased stability of the mRNA corresponding to the synthetic Env coding sequences versus the mRNA corresponding to the native Env coding sequences.

Modification of the Env polypeptide coding sequences resulted in improved expression relative to the wild-type
30 coding sequences in a number of mammalian cell lines. Similar Env polypeptide coding sequences can be obtained from a variety of isolates (families, sub-types, etc.). Env polypeptide encoding sequences derived from these variants can be optimized and tested for improved

expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 2).

5 2.1.5 **FURTHER MODIFICATION OF HIV-1 ENV NUCLEIC ACID
CODING SEQUENCES**

In addition to proteins containing HIV-related sequences, the Env encoding sequences of the present invention can be fused to other polypeptides (creating
10 chimeric polypeptides). Also, variations on the orientation of the Env and other coding sequences, relative to each other, are contemplated. Further, the HIV protein encoding cassettes of the present invention can be co-expressed using one vector or multiple vectors.
15 In addition, the polyproteins can be operably linked to the same or different promoters.

Env polypeptide coding sequences can be obtained from any HIV isolates (different families, subtypes, and strains) including but not limited to the isolates HIV_{11tb},
20 HIV_{SP2}, HIV_{us4}, HIV_{CM235}, HIV_{SP162}, HIV_{LAV}, HIV_{LAT}, HIV_{NI} (see, e.g., Myers et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos, New Mexico: Los Alamos National Laboratory). Synthetic
25 expression cassettes can be generated using such coding sequences as starting material by following the teachings of the present specification (e.g., see Example 1). Further, the synthetic expression cassettes (and purified polynucleotides) of the present invention include related
30 Env polypeptide coding sequences having greater than 90%, preferably greater than 92%, more preferably greater than 95%, and most preferably greater than 98% sequence identity to the synthetic expression cassette sequences disclosed herein (for example, SEQ ID NOs:71-72; and/or

the sequences presented in Tables 1A and 1B) when the sequences of the present invention are used as the query sequence.

5 2.1.6 **EXPRESSION OF SYNTHETIC SEQUENCES ENCODING HIV-1
 ENV AND RELATED POLYPEPTIDES**

Several synthetic Env-encoding sequences (expression
cassettes) of the present invention were cloned into a
number of different expression vectors (Example 1) to
10 evaluate levels of expression and production of Env
polypeptide. A modified synthetic coding sequence is
presented as synthetic Env expression cassettes (Example
1, e.g., Tables 1A and 1B). The synthetic DNA fragments
for Env were cloned into eucaryotic expression vectors
15 described in Example 1 and in Section 2.1.3 above,
including, a transient expression vector and CMV-
promoter-based mammalian vectors. Corresponding wild-
type sequences were cloned into the same vectors.

These vectors were then transfected into a several
20 different cell types, including a variety of mammalian
cell lines, (293, RD, COS-7, and CHO, cell lines
available, for example, from the A.T.C.C.). The cell
lines were cultured under appropriate conditions and the
levels of gp120, gp140 and gp160 Env expression in
25 supernatants were evaluated (Example 2). Env
polypeptides include, but are not limited to, for
example, native gp160, oligomeric gp140, monomeric gp120
as well as modified sequences of these polypeptides. The
results of these assays demonstrated that expression of
30 synthetic Env encoding sequences were significantly
higher than corresponding wild-type sequences (Example 2;
Tables 3 and 4).

Further, Western Blot analysis showed that cells
containing the synthetic Env expression cassette produced

the expected protein (gp120, gp140 or gp160) at higher per-cell concentrations than cells containing the native expression cassette. The Env proteins were seen in both cell lysates and supernatants. The levels of production were significantly higher in cell supernatants for cells transfected with the synthetic Env expression cassettes of the present invention as compared to wild type.

Fractionation of the supernatants from mammalian cells transfected with the synthetic Env expression cassettes showed that it provides superior production of Env proteins, relative to the wild-type Env sequences (Examples 2 and 3).

Efficient expression of these Env-containing polypeptides in mammalian cell lines provides the following benefits: the Env polypeptides are free of baculovirus or other viral contaminants; production by established methods approved by the FDA; increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Env-containing polypeptides in CHO cells which is less feasible in the absence of the increased expression obtained using the constructs of the present invention.

Exemplary cell lines (e.g., mammalian, yeast, insect, etc.) include those described above in Section 2.1.3 for Gag-containing constructs. Further, appropriate vectors and control elements (e.g., promoters, enhancers, polyadenylation sequences, etc.) for any given cell type can be selected, as described above in Section 2.1.3, by one having ordinary skill in the art in view of the teachings of the present specification and information known in the art about expression vectors. In addition, the recombinantly expressed polypeptides from synthetic Env-encoding expression cassettes are typically isolated and purified from lysed cells or culture media, as

described above for Gag-encoding expression cassettes. An exemplary purification is described in Example 4 and shown in Figure 60.

5 2.1.7 MODIFICATION OF HIV-1 TAT NUCLEIC ACID CODING
SEQUENCES

Another aspect of the present invention is the generation of HIV-1 tat protein coding sequences, and related sequences, having improved expression relative to
10 the corresponding wild-type sequence. Exemplary embodiments of the present invention are illustrated herein modifying the tat wild-type nucleotide sequence (SEQ ID NO:85, Figure 76) obtained from SF162 as described above. Exemplary synthetic tat constructs are
15 shown in SEQ ID NO:87, which depicts a tat construct encoding a full-length tat polypeptide from strain SF162; SEQ ID NO:88, which depicts a tat construct encoding a tat polypeptide having the cystein residue at position 22 changed; and SEQ ID NO:89, which depicts a tat construct
20 encoding the amino terminal portion of a tat polypeptide from strain SF162. The amino portion of the tat protein appears to contain many of the epitopes that induce an immune response. In addition, further modifications include replacement or deletion of the cystein residue at
25 position 22, for example with a valine residue, an alanine residue or a glycine residue (SEQ ID Nos: 88 and 89, Figures 79 and 81), see, e.g., Caputo et al. (1996) *Gene Ther.* 3:235. In Figure 81, which depicts a tat construct encoding the amino terminal portion of a tat
30 polypeptide, the nucleotides (nucleotides 64-66) encoding the cystein residues are underlined. The design and construction of suitable construct can be readily done using

the teachings of the present specification. As with Gag, pol, prot and Env, tat polypeptide coding sequences can be obtained from a variety of isolates (families, subtypes, etc.).

5 Modification of the tat polypeptide coding sequences result in improved expression relative to the wild-type coding sequences in a number of cell lines (e.g., mammalian, yeast, bacterial and insect cells). Tat polypeptide encoding sequences derived from these
10 variants can be optimized and tested for improved expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 2).

 Various forms of the different embodiments of the
15 invention, described herein, may be combined. For example, polynucleotides may be derived from the polynucleotide sequences of the present invention, including, but not limited to, coding sequences for Gag polypeptides, Env polypeptides, polymerase polypeptides,
20 protease polypeptides, tat polypeptides, and reverse transcriptase polypeptides. Further, the polynucleotide coding sequences of the present invention may be combined into multi-cistronic expression cassettes where typically each coding sequence for each polypeptide is preceded by
25 IRES sequences.

2.2 PRODUCTION OF VIRUS-LIKE PARTICLES AND USE OF THE CONSTRUCTS OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES

30 The group-specific antigens (Gag) of human immunodeficiency virus type-1 (HIV-1) self-assemble into noninfectious virus-like particles (VLP) that are released from various eucaryotic cells by budding (reviewed by Freed, E.O., *Virology* 251:1-15, 1998). The

synthetic expression cassettes of the present invention provide efficient means for the production of HIV-Gag virus-like particles (VLPs) using a variety of different cell types, including, but not limited to, mammalian
5 cells.

Viral particles can be used as a matrix for the proper presentation of an antigen entrapped or associated therewith to the immune system of the host. For example, U.S. Patent No. 4,722,840 describes hybrid particles
10 comprised of a particle-forming fragment of a structural protein from a virus, such as a particle-forming fragment of hepatitis B virus (HBV) surface antigen (HBsAg), fused to a heterologous polypeptide. Tindle et al., *Virology* (1994) 200:547-557, describes the production and use of
15 chimeric HBV core antigen particles containing epitopes of human papillomavirus (HPV) type 16 E7 transforming protein.

Adams et al., *Nature* (1987) 329:68-70, describes the recombinant production of hybrid HIVgp120:Ty VLPs in
20 yeast and Brown et al., *Virology* (1994) 198:477-488, the production of chimeric proteins consisting of the VP2 protein of human parvovirus B19 and epitopes from human herpes simplex virus type 1, as well as mouse hepatitis virus A59. Wagner et al., (*Virology* (1994) 200:162-175,
25 Brand et al., *J. Virol. Meth.* (1995) 51:153-168; *Virology* (1996) 220:128-140) and Wolf, et al., (EP 0 449 116 A1, published 2 October 1991; WO 96/30523, published 3
October 1996) describe the assembly of chimeric HIV-1 p55Gag particles. U.S. Patent No. 5,503,833 describes
30 the use of rotavirus VP6 spheres for encapsulating and delivering therapeutic agents.

2.2.1 VLP PRODUCTION USING THE SYNTHETIC EXPRESSION

CASSETTES OF THE PRESENT INVENTION

Experiments performed in support of the present invention have demonstrated that the synthetic expression cassettes of the present invention provide superior production of both protein and VLPs, relative to native coding sequences (Examples 7 and 15). Further, electron microscopic evaluation of VLP production (Examples 6 and 15, Figures 3A-B and 65A-F) showed that free and budding immature virus particles of the expected size were produced by cells containing the synthetic expression cassettes.

Using the synthetic expression cassettes of the present invention, rather than native coding sequences, for the production of virus-like particles provide several advantages. First, VLPs can be produced in enhanced quantity making isolation and purification of the VLPs easier. Second, VLPs can be produced in a variety of cell types using the synthetic expression cassettes, in particular, mammalian cell lines can be used for VLP production, for example, CHO cells. Production using CHO cells provides (i) VLP formation; (ii) correct myristylation and budding; (iii) absence of non-mammalian cell contaminants (e.g., insect viruses and/or cells); and (iv) ease of purification. The synthetic expression cassettes of the present invention are also useful for enhanced expression in cell-types other than mammalian cell lines. For example, infection of insect cells with baculovirus vectors encoding the synthetic expression cassettes resulted in higher levels of total protein yield and higher levels of VLP production (relative to wild-type coding sequences). Further, the final product from insect cells infected with the baculovirus-Gag synthetic expression cassettes

consistently contained lower amounts of contaminating insect proteins than the final product when wild-type coding sequences were used (Examples).

VLPs can spontaneously form when the particle-
5 forming polypeptide of interest is recombinantly expressed in an appropriate host cell. Thus, the VLPs produced using the synthetic expression cassettes of the present invention are conveniently prepared using recombinant techniques. As discussed below, the Gag
10 polypeptide encoding synthetic expression cassettes of the present invention can include other polypeptide coding sequences of interest (for example, Env, tat, rev, HIV protease, HIV polymerase, HCV core; see, Example 1). Expression of such synthetic expression cassettes yields
15 VLPs comprising the product of the synthetic expression cassette, as well as, the polypeptide of interest.

Once coding sequences for the desired particle-forming polypeptides have been isolated or synthesized, they can be cloned into any suitable vector or replicon
20 for expression. Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. See, generally, Ausubel et al, *supra* or Sambrook et al, *supra*. The vector is then used to transform an appropriate host
25 cell. Suitable recombinant expression systems include, but are not limited to, bacterial, mammalian, baculovirus/insect, vaccinia, Semliki Forest virus (SFV), Alphaviruses (such as, Sindbis, Venezuelan Equine Encephalitis (VEE)), mammalian, yeast and Xenopus
30 expression systems, well known in the art. Particularly preferred expression systems are mammalian cell lines, vaccinia, Sindbis, insect and yeast systems.

For example, a number of mammalian cell lines are known in the art and include immortalized cell lines

- available from the American Type Culture Collection (A.T.C.C.), such as, but not limited to, Chinese hamster ovary (CHO) cells, 293 cells, HeLa cells, baby hamster kidney (BHK) cells, mouse myeloma (SB20), monkey kidney cells (COS), as well as others. Similarly, bacterial hosts such as *E. coli*, *Bacillus subtilis*, and *Streptococcus spp.*, will find use with the present expression constructs. Yeast hosts useful in the present invention include *inter alia*, *Saccharomyces cerevisiae*, *Candida albicans*, *Candida maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guilliermondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Insect cells for use with baculovirus expression vectors include, *inter alia*, *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni*. See, e.g., Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987). Fungal hosts include, for example, *Aspergillus*.
- Viral vectors can be used for the production of particles in eucaryotic cells, such as those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. Additionally, a vaccinia based infection/transfection system, as described in Tomei et al., *J. Virol.* (1993) 67:4017-4026 and Selby et al., *J. Gen. Virol.* (1993) 74:1103-1113, will also find use with the present invention. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the DNA of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus

recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. Alternately, T7 can be added as a purified protein or enzyme as in the "Progenitor" system (Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130). The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation product(s).

Depending on the expression system and host selected, the VLPs are produced by growing host cells transformed by an expression vector under conditions whereby the particle-forming polypeptide is expressed and VLPs can be formed. The selection of the appropriate growth conditions is within the skill of the art. If the VLPs are formed intracellularly, the cells are then disrupted, using chemical, physical or mechanical means, which lyse the cells yet keep the VLPs substantially intact. Such methods are known to those of skill in the art and are described in, e.g., *Protein Purification Applications: A Practical Approach*, (E.L.V. Harris and S. Angal, Eds., 1990).

The particles are then isolated (or substantially purified) using methods that preserve the integrity thereof, such as, by density gradient centrifugation, e.g., sucrose gradients, PEG-precipitation, pelleting, and the like (see, e.g., Kirnbauer et al. *J. Virol.* (1993) 67:6929-6936), as well as standard purification techniques including, e.g., ion exchange and gel filtration chromatography.

VLPs produced by cells containing the synthetic expression cassettes of the present invention can be used to elicit an immune response when administered to a subject. One advantage of the present invention is that VLPs can be produced by mammalian cells carrying the

synthetic expression cassettes at levels previously not possible. As discussed above, the VLPs can comprise a variety of antigens in addition to the Gag polypeptides (e.g., Env, tat, Gag-protease, Gag-polymerase, Gag-HCV-core). Purified VLPs, produced using the synthetic expression cassettes of the present invention, can be administered to a vertebrate subject, usually in the form of vaccine compositions. Combination vaccines may also be used, where such vaccines contain, for example, other subunit proteins derived from HIV or other organisms (e.g., env) or gene delivery vaccines encoding such antigens. Administration can take place using the VLPs formulated alone or formulated with other antigens. Further, the VLPs can be administered prior to, concurrent with, or subsequent to, delivery of the synthetic expression cassettes for DNA immunization (see below) and/or delivery of other vaccines. Also, the site of VLP administration may be the same or different as other vaccine compositions that are being administered. Gene delivery can be accomplished by a number of methods including, but are not limited to, immunization with DNA, alphavirus vectors, pox virus vectors, and vaccinia virus vectors.

VLP immune-stimulating (or vaccine) compositions can include various excipients, adjuvants, carriers, auxiliary substances, modulating agents, and the like. The immune stimulating compositions will include an amount of the VLP/antigen sufficient to mount an immunological response. An appropriate effective amount can be determined by one of skill in the art. Such an amount will fall in a relatively broad range that can be determined through routine trials and will generally be an amount on the order of about 0.1 μ g to about 1000 μ g,

more preferably about 1 μg to about 300 μg , of VLP/antigen.

A carrier is optionally present which is a molecule that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP, et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen may be conjugated to a bacterial toxoid, such as toxoid from diphtheria, tetanus, cholera, etc., as well as toxins derived from *E. coli*.

Such adjuvants include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (International Publication No. WO 90/14837), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated

into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) 5 either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group 10 consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particle generated therefrom such as 15 ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, etc.), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), beta chemokines (MIP, 1- 20 alpha, 1-beta Rantes, etc.); (6) detoxified mutants of a bacterial ADP-ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an *E. coli* heat-labile toxin (LT), particularly LT-K63 (where lysine is substituted for the wild-type amino acid at position 63) 25 LT-R72 (where arginine is substituted for the wild-type amino acid at position 72), CT-S109 (where serine is substituted for the wild-type amino acid at position 109), and PT-K9/G129 (where lysine is substituted for the wild-type amino acid at position 9 and glycine 30 substituted at position 129) (see, e.g., International Publication Nos. W093/13202 and W092/19265); and (7)

other substances that act as immunostimulating agents to enhance the effectiveness of the composition.

Muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acteyl-normuramyl-L-alanyl-D-isogluatme (nor-MDP), N-acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

Dosage treatment with the VLP composition may be a single dose schedule or a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals, chosen to maintain and/or reinforce the immune response, for example at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the potency of the modality, the vaccine delivery employed, the need of the subject and be dependent on the judgment of the practitioner.

If prevention of disease is desired (e.g., reduction of symptoms, recurrences or of disease progression), the antigen carrying VLPs are generally administered prior to primary infection with the pathogen of interest. If treatment is desired, e.g., the reduction of symptoms or recurrences, the VLP compositions are generally administered subsequent to primary infection.

2.2.2 USING THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES

A number of viral based systems have been developed for use as gene transfer vectors for mammalian host cells. For example, retroviruses (in particular,

lentiviral vectors) provide a convenient platform for gene delivery systems. A coding sequence of interest (for example, a sequence useful for gene therapy applications) can be inserted into a gene delivery vector and packaged in retroviral particles using techniques known in the art. Recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described, including, for example, the following: (U.S. Patent No. 5,219,740; Miller et al. (1989) *Biotechniques* 7:980; Miller, A.D. (1990) *Human Gene Therapy* 1:5; Scarpa et al. (1991) *Virology* 180:849; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033; Boris-Lawrie et al. (1993) *Cur. Opin. Genet. Develop.* 3:102; GB 2200651; EP 0415731; EP 0345242; WO 89/02468; WO 89/05349; WO 89/09271; WO 90/02806; WO 90/07936; WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; WO 93/11230; WO 93/10218; WO 91/02805; in U.S. 5,219,740; U.S. 4,405,712; U.S. 4,861,719; U.S. 4,980,289 and U.S. 4,777,127; in U.S. Serial No. 07/800,921; and in Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53:83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci USA* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Sequences useful for gene therapy applications include, but are not limited to, the following. Factor VIII cDNA, including derivatives and deletions thereof (International Publication Nos. WO 96/21035, WO 97/03193, WO 97/03194, WO 97/03195, and WO 97/03191). Factor IX cDNA (Kurachi et al. (1982) *Proc. Natl. Acad. Sci. USA* 79:6461-6464). Factor V cDNA can be obtained from pMT2-V (Jenny (1987) *Proc. Natl. Acad. Sci. USA* 84:4846, A.T.C.C. Deposit No. 40515). A full-length factor V

cDNA, or a B domain deletion or B domain substitution thereof, can be used. B domain deletions of factor V, include those reported by Marquette (1995) *Blood* 86:3026 and Kane (1990) *Biochemistry* 29:6762. Antithrombin III
5 cDNA (Prochownik (1983) *J. Biol. Chem.* 258:8389, A.T.C.C. Deposit No. 57224/57225). Protein C encoding cDNA (Foster (1984) *Proc. Natl. Acad. Sci. USA* 81:4766; Beckmann (1985) *Nucleic Acids Res.* 13:5233). Prothrombin cDNA can be obtained by restriction enzyme digestion of a
10 published vector (Degen (1983) *Biochemistry* 22:2087). The endothelial cell surface protein, thrombomodulin, is a necessary cofactor for the normal activation of protein C by thrombin. A soluble recombinant form has been described (Parkinson (1990) *J. Biol. Chem.* 265:12602;
15 Jackman (1987) *Proc. Natl. Acad. Sci. USA* 84:6425; Shirai (1988) *J. Biochem.* 103:281; Wen (1987) *Biochemistry* 26:4350; Suzuki (1987) *EMBO J.* 6:1891, A.T.C.C. Deposit No. 61348, 61349).

Many genetic diseases caused by inheritance of
20 defective genes result in the failure to produce normal gene products, for example, thalassemia, phenylketonuria, Lesch-Nyhan syndrome, severe combined immunodeficiency (SCID), hemophilia A and B, cystic fibrosis, Duchenne's Muscular Dystrophy, inherited emphysema and familial
25 hypercholesterolemia (Mulligan et al. (1993) *Science* 260:926; Anderson et al. (1992) *Science* 256:808; Friedman et al. (1989) *Science* 244:1275). Although genetic diseases may result in the absence of a gene product, endocrine disorders, such as diabetes and
30 hypopituitarism, are caused by the inability of the gene to produce adequate levels of the appropriate hormone insulin and human growth hormone respectively.

In one aspect, gene therapy employing the constructs and methods of the present invention involves the

introduction of normal recombinant genes into T cells so that new or missing proteins are produced by the T cells after introduction or reintroduction thereof into a patient. A number of genetic diseases have been selected for treatment with gene therapy, including adenine deaminase deficiency, cystic fibrosis, α_1 -antitrypsin deficiency, Gaucher's syndrome, as well as non-genetic diseases.

In particular, Gaucher's syndrome is a genetic disorder characterized by a deficiency of the enzyme glucocerebrosidase. This enzyme deficiency leads to the accumulation of glucocerebroside in the lysosomes of all cells in the body. For a review see *Science* 256:794 (1992) and Scriver et al., *The Metabolic Basis of Inherited Disease*, 6th ed., vol. 2, page 1677). Thus, gene transfer vectors that express glucocerebrosidase can be constructed for use in the treatment of this disorder. Likewise, gene transfer vectors encoding lactase can be used in the treatment of hereditary lactose intolerance, those expressing AD can be used for treatment of ADA deficiency, and gene transfer vectors encoding α_1 -antitrypsin can be used to treat α_1 -antitrypsin deficiency. See Ledley, F.D. (1987) *J. Pediatrics* 110:157-174, Verma, I. (Nov. 1987) *Scientific American* pp. 68-84, and International Publication No. WO 95/27512 entitled "Gene Therapy Treatment for a Variety of Diseases and Disorders," for a description of gene therapy treatment of genetic diseases.

In still further embodiments of the invention, nucleotide sequences which can be incorporated into a gene transfer vector include, but are not limited to, proteins associated with enzyme-deficiency disorders, such as the cystic fibrosis transmembrane regulator (see, for example, U.S. Patent No. 5,240,846 and Larrick et al.

(1991) *Gene Therapy Applications of Molecular Biology*, Elsevier, New York and adenosine deaminase (ADA) (see U.S. Patent No. 5,399,346); growth factors, or an agonist or antagonist of a growth factor (Bandara et al. (1992) *DNA and Cell Biology*, 11:227); one or more tumor suppressor genes such as p53, Rb, or C-CAM1 (Kleinerman et al. (1995) *Cancer Research* 55:2831); a molecule that modulates the immune system of an organism, such as a HLA molecule (Nabel et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:11307); a ribozyme (Larsson et al. (1996) *Virology* 219:161); a peptide nucleic acid (Hirshman et al. (1996) *J. Invest. Med.* 44:347); an antisense molecule (Bordier et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:9383) which can be used to down-regulate the expression or synthesis of aberrant or foreign proteins, such as HIV proteins or a wide variety of oncogenes such as p53 (Hesketh, *The Oncogene Facts Book*, Academic Press, New York, (1995); a biopharmaceutical agent or antisense molecule used to treat HIV-infection, such as an inhibitor of p24 (Nakashima et al. (1994) *Nucleic Acids Res.* 22:5004); or reverse-transcriptase (see, Bordier, *supra*).

Other proteins of therapeutic interest can be expressed *in vivo* by gene transfer vectors using the methods of the invention. For instance sustained *in vivo* expression of tissue factor inhibitory protein (TFPI) is useful for treatment of conditions including sepsis and DIC and in preventing reperfusion injury. (See International Publications Nos. WO 93/24143, WO 93/25230 and WO 96/06637). Nucleic acid sequences encoding various forms of TFPI can be obtained, for example, as described in US Patent Nos. 4,966,852; 5,106,833; and 5,466,783, and incorporated into the gene transfer vectors described herein.

Erythropoietin (EPO) and leptin can also be expressed *in vivo* from genetically modified T cells according to the methods of the invention. For instance EPO is useful in gene therapy treatment of a variety of disorders including anemia (see International Publication No. WO 95/13376 entitled "Gene Therapy for Treatment of Anemia"). Sustained delivery of leptin by the methods of the invention is useful in treatment of obesity. See International Publication No. WO 96/05309 for a description of the leptin gene and the use thereof in the treatment of obesity.

A variety of other disorders can also be treated by the methods of the invention. For example, sustained *in vivo* systemic production of apolipoprotein E or apolipoprotein A from genetically modified T cells can be used for treatment of hyperlipidemia (see Breslow et al. (1994) *Biotechnology* 12:365). Sustained production of angiotensin receptor inhibitor (Goodfriend et al. (1996) *N. Engl. J. Med.* 334:1469) can be provided by the methods described herein. As yet an additional example, the long term *in vivo* systemic production of angiostatin is useful in the treatment of a variety of tumors. (See O'Reilly et al. (1996) *Nature Med.* 2:689).

In other embodiments, gene transfer vectors can be constructed to encode a cytokine or other immunomodulatory molecule. For example, nucleic acid sequences encoding native IL-2 and gamma-interferon can be obtained as described in US Patent Nos. 4,738,927 and 5,326,859, respectively, while useful muteins of these proteins can be obtained as described in U.S. Patent No. 4,853,332. Nucleic acid sequences encoding the short and long forms of mCSF can be obtained as described in US Patent Nos. 4,847,201 and 4,879,227, respectively. In particular aspects of the invention, retroviral vectors

expressing cytokine or immunomodulatory genes can be produced as described herein (for example, employing the packaging cell lines of the present invention) and in International Application No. PCT US 94/02951, entitled
5 "Compositions and Methods for Cancer Immunotherapy."

- Examples of suitable immunomodulatory molecules for use herein include the following: IL-1 and IL-2 (Karupiah et al. (1990) *J. Immunology* 144:290-298, Weber et al. (1987) *J. Exp. Med.* 166:1716-1733, Gansbacher et al. (1990) *J. Exp. Med.* 172:1217-1224, and U.S. Patent No. 4,738,927); IL-3 and IL-4 (Tepper et al. (1989) *Cell* 57:503-512, Golumbek et al. (1991) *Science* 254:713-716, and U.S. Patent No. 5,017,691); IL-5 and IL-6 (Brakenhof et al. (1987) *J. Immunol.* 139:4116-4121, and
15 International Publication No. WO 90/06370); IL-7 (U.S. Patent No. 4,965,195); IL-8, IL-9, IL-10, IL-11, IL-12, and IL-13 (*Cytokine Bulletin*, Summer 1994); IL-14 and IL-15; alpha interferon (Pinter et al. (1991) *Drugs* 42:749-765, U.S. Patent Nos. 4,892,743 and 4,966,843,
20 International Publication No. WO 85/02862, Nagata et al. (1980) *Nature* 284:316-320, Familletti et al. (1981) *Methods in Enz.* 78:387-394, Twu et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2046-2050, and Faktor et al. (1990) *Oncogene* 5:867-872); beta-interferon (Seif et al. (1991)
25 *J. Virol.* 65:664-671); gamma-interferons (Radford et al. (1991) *The American Society of Hepatology* 20082015, Watanabe et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:9456-9460, Gansbacher et al. (1990) *Cancer Research* 50:7820-7825, Maio et al. (1989) *Can. Immunol.*
30 *Immunother.* 30:34-42, and U.S. Patent Nos. 4,762,791 and 4,727,138); G-CSF (U.S. Patent Nos. 4,999,291 and 4,810,643); GM-CSF (International Publication No. WO 85/04188); tumor necrosis factors (TNFs) (Jayaraman et al. (1990) *J. Immunology* 144:942-951); CD3 (Krissanen et

al. (1987) *Immunogenetics* 26:258-266); ICAM-1 (Altman et al. (1989) *Nature* 338:512-514, Simmons et al. (1988) *Nature* 331:624-627); ICAM-2, LFA-1, LFA-3 (Wallner et al. (1987) *J. Exp. Med.* 166:923-932); MHC class I molecules, 5 MHC class II molecules, B7.1-.3, β_2 -microglobulin (Parnes et al. (1981) *Proc. Natl. Acad. Sci. USA* 78:2253-2257); chaperones such as calnexin; and MHC-linked transporter proteins or analogs thereof (Powis et al. (1991) *Nature* 354:528-531). Immunomodulatory factors may also be 10 agonists, antagonists, or ligands for these molecules. For example, soluble forms of receptors can often behave as antagonists for these types of factors, as can mutated forms of the factors themselves.

Nucleic acid molecules that encode the above- 15 described substances, as well as other nucleic acid molecules that are advantageous for use within the present invention, may be readily obtained from a variety of sources, including, for example, depositories such as the American Type Culture Collection, or from commercial 20 sources such as British Bio-Technology Limited (Cowley, Oxford England). Representative examples include BBG 12 (containing the GM-CSF gene coding for the mature protein of 127 amino acids), BBG 6 (which contains sequences encoding gamma interferon), A.T.C.C. Deposit No. 39656 25 (which contains sequences encoding TNF), A.T.C.C. Deposit No. 20663 (which contains sequences encoding alpha-interferon), A.T.C.C. Deposit Nos. 31902, 31902 and 39517 (which contain sequences encoding beta-interferon), A.T.C.C. Deposit No. 67024 (which contains a sequence 30 which encodes Interleukin-1b), A.T.C.C. Deposit Nos. 39405, 39452, 39516, 39626 and 39673 (which contain sequences encoding Interleukin-2), A.T.C.C. Deposit Nos. 59399, 59398, and 67326 (which contain sequences encoding Interleukin-3), A.T.C.C. Deposit No. 57592 (which

contains sequences encoding Interleukin-4), A.T.C.C. Deposit Nos. 59394 and 59395 (which contain sequences encoding Interleukin-5), and A.T.C.C. Deposit No. 67153 (which contains sequences encoding Interleukin-6).

5 Plasmids containing cytokine genes or immunomodulatory genes (International Publication Nos. WO 94/02951 and WO 96/21015) can be digested with appropriate restriction enzymes, and DNA fragments containing the particular gene of interest can be inserted into a gene
10 transfer vector using standard molecular biology techniques. (See, e.g., Sambrook et al., *supra.*, or Ausubel et al. (eds) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience).

Exemplary hormones, growth factors and other
15 proteins which are useful for long term expression are described, for example, in European Publication No. 0437478B1, entitled "Cyclodextrin-Peptide Complexes." Nucleic acid sequences encoding a variety of hormones can be used, including those encoding human growth hormone,
20 insulin, calcitonin, prolactin, follicle stimulating hormone (FSH), luteinizing hormone (LH), human chorionic gonadotropin (HCG), and thyroid stimulating hormone (TSH). A variety of different forms of IGF-1 and IGF-2 growth factor polypeptides are also well known the art
25 and can be incorporated into gene transfer vectors for long term expression *in vivo*. See, e.g., European Patent No. 0123228B1, published for grant September 19, 1993, entitled "Hybrid DNA Synthesis of Mature Insulin-like Growth Factors." As an additional example, the long term
30 *in vivo* expression of different forms of fibroblast growth factor can also be effected employing the compositions and methods of invention. See, e.g., U.S. Patent Nos. 5,464,774, 5,155,214, and 4,994,559 for a description of different fibroblast growth factors.

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene
5 from a vector known to include the same. For example, plasmids which contain sequences that encode altered cellular products may be obtained from a depository such as the A.T.C.C., or from commercial sources. Plasmids containing the nucleotide sequences of interest can be
10 digested with appropriate restriction enzymes, and DNA fragments containing the nucleotide sequences can be inserted into a gene transfer vector using standard molecular biology techniques.

Alternatively, cDNA sequences for use with the
15 present invention may be obtained from cells which express or contain the sequences, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain and isolate DNA.
20 Briefly, mRNA from a cell which expresses the gene of interest can be reverse transcribed with reverse transcriptase using oligo-dT or random primers. The single stranded cDNA may then be amplified by PCR (see U.S. Patent Nos. 4,683,202, 4,683,195 and 4,800,159, see
25 also *PCR Technology: Principles and Applications for DNA Amplification*, Erlich (ed.), Stockton Press, 1989)) using oligonucleotide primers complementary to sequences on either side of desired sequences.

The nucleotide sequence of interest can also be
30 produced synthetically, rather than cloned, using a DNA synthesizer (e.g., an Applied Biosystems Model 392 DNA Synthesizer, available from ABI, Foster City, California). The nucleotide sequence can be designed with the appropriate codons for the expression product

desired. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge (1981) *Nature* 292:756; Nambair et al. (1984) *Science* 223:1299; Jay et al. (1984) *J. Biol. Chem.* 259:6311.

The synthetic expression cassettes of the present invention can be employed in the construction of packaging cell lines for use with retroviral vectors.

One type of retrovirus, the murine leukemia virus, or "MLV", has been widely utilized for gene therapy applications (see generally Mann et al. (*Cell* 33:153, 1993), Cane and Mulligan (*Proc. Nat'l. Acad. Sci. USA* 81:6349, 1984), and Miller et al., *Human Gene Therapy* 1:5-14, 1990).

Lentiviral vectors typically, comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, a promoter operably linked to one or more genes of interest, an origin of second strand DNA synthesis and a 3' lentiviral LTR, wherein the lentiviral vector contains a nuclear transport element. The nuclear transport element may be located either upstream (5') or downstream (3') of a coding sequence of interest. Within certain embodiments, the nuclear transport element is not RRE. Within one embodiment the packaging signal is an extended packaging signal. Within other embodiments the promoter is a tissue specific promoter, or, alternatively, a promoter such as CMV. Within other embodiments, the lentiviral vector further comprises an internal ribosome entry site.

A wide variety of lentiviruses may be utilized within the context of the present invention, including for example, lentiviruses selected from the group consisting of HIV, HIV-1, HIV-2, FIV and SIV.

In one embodiment of the present invention synthetic Env and/or Gag-polymerase expression cassettes are provided comprising a promoter and a sequence encoding synthetic Gag-polymerase (SEQ ID NO:6) and at least one
5 of vpr, vpu, nef or vif, wherein the promoter is operably linked to Gag-polymerase and vpr, vpu, nef or vif.

Within yet another aspect of the invention, host cells (e.g., packaging cell lines) are provided which contain any of the expression cassettes described herein.
10 For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic Env and/or Gag-polymerase, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding Env and/or Gag-polymerase. Packaging cell lines may further comprise a
15 promoter and a sequence encoding tat, rev, or an envelope, wherein the promoter is operably linked to the sequence encoding tat, rev, or, the envelope. The packaging cell line may further comprise a sequence
20 encoding any one or more of nef, vif, vpu or vpr.

In one embodiment, the expression cassette (carrying, for example, the synthetic Env, synthetic tat and/or synthetic Gag-polymerase) is stably integrated. The packaging cell line, upon introduction of a
25 lentiviral vector, typically produces viral particles. The promoter regulating expression of the synthetic expression cassette may be inducible. Typically, the packaging cell line, upon introduction of a lentiviral vector, produces viral particles that are essentially
30 free of replication competent virus.

Packaging cell lines are provided comprising an expression cassette which directs the expression of a synthetic Env (or Gag-polymerase) gene, an expression cassette which directs the expression of a Gag (or Env)

gene optimized for expression (e.g., Andre, S., et al.,
Journal of Virology 72(2):1497-1503, 1998; Haas, J., et
al., *Current Biology* 6(3):315-324, 1996). A lentiviral
vector is introduced into the packaging cell line to
5 produce a vector particle producing cell line.

As noted above, lentiviral vectors can be designed
to carry or express a selected gene(s) or sequences of
interest. Lentiviral vectors may be readily constructed
from a wide variety of lentiviruses (see RNA Tumor
10 Viruses, Second Edition, Cold Spring Harbor Laboratory,
1985). Representative examples of lentiviruses included
HIV, HIV-1, HIV-2, FIV and SIV. Such lentiviruses may
either be obtained from patient isolates, or, more
preferably, from depositories or collections such as the
15 American Type Culture Collection, or isolated from known
sources using available techniques.

Portions of the lentiviral gene delivery vectors (or
vehicles) may be derived from different viruses. For
example, in a given recombinant lentiviral vector, LTRs
20 may be derived from an HIV, a packaging signal from SIV,
and an origin of second strand synthesis from HrV-2.
Lentiviral vector constructs may comprise a 5' lentiviral
LTR, a tRNA binding site, a packaging signal, one or more
heterologous sequences, an origin of second strand DNA
25 synthesis and a 3' LTR, wherein said lentiviral vector
contains a nuclear transport element that is not RRE.

Briefly, Long Terminal Repeats ("LTRs") are
subdivided into three elements, designated U5, R and U3.
These elements contain a variety of signals which are
30 responsible for the biological activity of a retrovirus,
including for example, promoter and enhancer elements
which are located within U3. LTRs may be readily
identified in the provirus (integrated DNA form) due to
their precise duplication at either end of the genome.

As utilized herein, a 5' LTR should be understood to include a 5' promoter element and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector. The 3' LTR should be understood to include a polyadenylation signal, and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector.

The tRNA binding site and origin of second strand DNA synthesis are also important for a retrovirus to be biologically active, and may be readily identified by one of skill in the art. For example, retroviral tRNA binds to a tRNA binding site by Watson-Crick base pairing, and is carried with the retrovirus genome into a viral particle. The tRNA is then utilized as a primer for DNA synthesis by reverse transcriptase. The tRNA binding site may be readily identified based upon its location just downstream from the 5'LTR. Similarly, the origin of second strand DNA synthesis is, as its name implies, important for the second strand DNA synthesis of a retrovirus. This region, which is also referred to as the poly-purine tract, is located just upstream of the 3'LTR.

In addition to a 5' and 3' LTR, tRNA binding site, and origin of second strand DNA synthesis, recombinant retroviral vector constructs may also comprise a packaging signal, as well as one or more genes or coding sequences of interest. In addition, the lentiviral vectors have a nuclear transport element which, in preferred embodiments is not RRE. Representative examples of suitable nuclear transport elements include the element in Rous sarcoma virus (Ogert, et al., *J ViroL* 70, 3834-3843, 1996), the element in Rous sarcoma virus (Liu & Mertz, *Genes & Dev.*, 9, 1766-1789, 1995) and the element in the genome of simian retrovirus type I

(Zolotukhin, et al., *J Virol.* 68, 7944-7952, 1994).

Other potential elements include the elements in the histone gene (Kedes, *Annu. Rev. Biochem.* 48, 837-870, 1970), the α -interferon gene (Nagata et al., *Nature* 287, 401-408, 1980), the β -adrenergic receptor gene (Koilkka, et al., *Nature* 329, 75-79, 1987), and the c-Jun gene (Hattorie, et al., *Proc. Natl. Acad. Sci. USA* 85, 9148-9152, 1988).

Recombinant lentiviral vector constructs typically lack both *Gag-polymerase* and *env* coding sequences. Recombinant lentiviral vector typically contain less than 20, preferably 15, more preferably 10, and most preferably 8 consecutive nucleotides found in *Gag-polymerase* or *env* genes. One advantage of the present invention is that the synthetic *Gag-polymerase* expression cassettes, which can be used to construct packaging cell lines for the recombinant retroviral vector constructs, have little homology to wild-type *Gag-polymerase* sequences and thus considerably reduce or eliminate the possibility of homologous recombination between the synthetic and wild-type sequences.

Lentiviral vectors may also include tissue-specific promoters to drive expression of one or more genes or sequences of interest. For example, lentiviral vector particles of the invention can contain a liver specific promoter to maximize the potential for liver specific expression of the exogenous DNA sequence contained in the vectors. Preferred liver specific promoters include the hepatitis B X-gene promoter and the hepatitis B core protein promoter. These liver specific promoters are preferably employed with their respective enhancers. The enhancer element can be linked at either the 5' or the 3' end of the nucleic acid encoding the sequences of interest. The hepatitis B X gene promoter and its

enhancer can be obtained from the viral genome as a 332 base pair *EcoRV-NcoI* DNA fragment employing the methods described in Twu, et al., *J Virol.* 61:3448-3453, 1987. The hepatitis B core protein promoter can be obtained
5 from the viral genome as a 584 base pair *BamHI-BglII* DNA fragment employing the methods described in Gerlach, et al., *Virology* 189:59-66, 1992. It may be necessary to remove the negative regulatory sequence in the *BamHI-BglII* fragment prior to inserting it. Other liver
10 specific promoters include the AFP (alpha fetal protein) gene promoter and the albumin gene promoter, as disclosed in EP Patent Publication 0 415 731, the -1 antitrypsin gene promoter, as disclosed in Rettenger, et al., *Proc. Natl. Acad. Sci.* 91:1460-1464, 1994, the fibrinogen
15 gene promoter, the APO-A1 (Apolipoprotein A1) gene promoter, and the promoter genes for liver transference enzymes such as, for example, SGOT, SGPT and glutamyl transferase. See also PCT Patent Publications WO 90/07936 and WO 91/02805 for a description of the use of
20 liver specific promoters in lentiviral vector particles.

Lentiviral vector constructs may be generated such that more than one gene of interest is expressed. This may be accomplished through the use of di- or oligo-cistronic cassettes (e.g., where the coding regions are
25 separated by 80 nucleotides or less, see generally Levin et al., *Gene* 108:167-174, 1991), or through the use of Internal Ribosome Entry Sites ("IRES").

Packaging cell lines suitable for use with the above described recombinant retroviral vector constructs may be
30 readily prepared given the disclosure provided herein. Briefly, the parent cell line from which the packaging cell line is derived can be selected from a variety of

mammalian cell lines, including for example, 293, RD, COS-7, CHO, BHK, VERO, HT1080, and myeloma cells.

After selection of a suitable host cell for the generation of a packaging cell line, one or more
5 expression cassettes are introduced into the cell line in order to complement or supply in trans components of the vector which have been deleted.

Representative examples of suitable expression cassettes have been described herein and include
10 synthetic Env, tat, Gag, synthetic Gag-protease, synthetic Gag-reverse transcriptase and synthetic Gag-polymerase expression cassettes, which comprise a promoter and a sequence encoding, e.g., Env, tat, or Gag-polymerase and at least one of vpr, vpu, nef or vif,
15 wherein the promoter is operably linked to Env, tat or Gag-polymerase and vpr, vpu, nef or vif. As described above, optimized Env, Gag and/or tat coding sequences may also be utilized in various combinations in the generation of packaging cell lines.

Utilizing the above-described expression cassettes, a wide variety of packaging cell lines can be generated. For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic HIV (e.g., Gag, Env, tat,
25 Gag-polymerase, Gag-reverse transcriptase or Gag-protease) polypeptide, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding the HIV polypeptide. Within other aspects, packaging cell lines are provided comprising a promoter
30 and a sequence encoding Gag, tat, rev, or an envelope (e.g., HIV env), wherein the promoter is operably linked to the sequence encoding Gag, tat, rev, or, the envelope. Within further embodiments, the packaging cell line may comprise a sequence encoding any one or more of nef, vif,

vpu or vpr. For example, the packaging cell line may contain only nef, vif, vpu, or vpr alone, nef and vif, nef and vpu, nef and vpr, vif and vpu, vif and vpr, vpu and vpr, nef vif and vpu, nef vif and vpr, nef vpu and vpr, vpr, vpr vpu and vpr, or, all four of nef vif vpu and vpr.

In one embodiment, the expression cassette is stably integrated. Within another embodiment, the packaging cell line, upon introduction of a lentiviral vector, produces particles. Within further embodiments the promoter is inducible. Within certain preferred embodiments of the invention, the packaging cell line, upon introduction of a lentiviral vector, produces particles that are free of replication competent virus.

The synthetic cassettes containing optimized coding sequences are transfected into a selected cell line. Transfected cells are selected that (i) carry, typically, integrated, stable copies of the Gag, Pol, and Env coding sequences, and (ii) are expressing acceptable levels of these polypeptides (expression can be evaluated by methods known in the prior art, e.g., see Examples 1-4). The ability of the cell line to produce VLPs may also be verified (Examples 6, 7 and 15).

A sequence of interest is constructed into a suitable viral vector as discussed above. This defective virus is then transfected into the packaging cell line. The packaging cell line provides the viral functions necessary for producing virus-like particles into which the defective viral genome, containing the sequence of interest, are packaged. These VLPs are then isolated and can be used, for example, in gene delivery or gene therapy.

Further, such packaging cell lines can also be used to produce VLPs alone, which can, for example, be used as

adjuvants for administration with other antigens or in vaccine compositions. Also, co-expression of a selected sequence of interest encoding a polypeptide (for example, an antigen) in the packaging cell line can also result in the entrapment and/or association of the selected polypeptide in/with the VLPs.

2.3 DNA IMMUNIZATION AND GENE DELIVERY

A variety of polypeptide antigens can be used in the practice of the present invention. Polypeptide antigens can be included in DNA immunization constructs containing, for example, any of the synthetic expression cassettes described herein fused in-frame to a coding sequence for the polypeptide antigen, where expression of the construct results in VLPs presenting the antigen of interest. Antigens can be derived from a wide variety of viruses, bacteria, fungi, plants, protozoans and other parasites. For example, the present invention will find use for stimulating an immune response against a wide variety of proteins from the herpesvirus family, including proteins derived from herpes simplex virus (HSV) types 1 and 2, such as HSV-1 and HSV-2 gB, gD, gH, VP16 and VP22; antigens derived from varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus (CMV) including CMV gB and gH; and antigens derived from other human herpesviruses such as HHV6 and HHV7. (See, e.g. Chee et al., *Cytomegaloviruses* (J.K. McDougall, ed., Springer-Verlag 1990) pp. 125-169, for a review of the protein coding content of cytomegalovirus; McGeoch et al., *J. Gen. Virol.* (1988) 69:1531-1574, for a discussion of the various HSV-1 encoded proteins; U.S. Patent No. 5,171,568 for a discussion of HSV-1 and HSV-2 gB and gD proteins and the genes encoding therefore; Baer et al., *Nature* (1984) 310:207-211, for the identification of

protein coding sequences in an EBV genome; and Davison and Scott, *J. Gen. Virol.* (1986) 67:1759-1816, for a review of VZV.)

5 Additionally, immune responses to antigens from the hepatitis family of viruses, including hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the delta hepatitis virus (HDV), hepatitis E virus (HEV), and hepatitis G virus, can also be stimulated using the constructs of the present invention. By way of example, 10 the HCV genome encodes several viral proteins, including E1 (also known as E) and E2 (also known as E2/NSI), which will find use with the present invention (see, Houghton et al. *Hepatology* (1991) 14:381-388, for a discussion of HCV proteins, including E1 and E2). The δ -antigen from 15 HDV can also be used (see, e.g., U.S. Patent No. 5,389,528, for a description of the δ -antigen).

 Similarly, influenza virus is another example of a virus for which the present invention will be particularly useful. Specifically, the envelope 20 glycoproteins HA and NA of influenza A are of particular interest for generating an immune response. Numerous HA subtypes of influenza A have been identified (Kawaoka et al., *Virology* (1990) 179:759-767; Webster et al. "Antigenic variation among type A influenza viruses," p. 25 127-168. In: P. Palese and D.W. Kingsbury (ed.), *Genetics of influenza viruses*. Springer-Verlag, New York).

 Other antigens of particular interest to be used in the practice of the present invention include antigens and polypeptides derived therefrom from human 30 papillomavirus (HPV), such as one or more of the various early proteins including E6 and E7; tick-borne encephalitis viruses; and HIV-1 (also known as HTLV-III, LAV, ARV, etc.), including, but not limited to, antigens such as gp120, gp41, gp160, Gag and pol from a variety of

- isolates including, but not limited to, HIV_{111b}, HIV_{BF2}, HIV-1_{SP162}, HIV-1_{SP170}, HIV_{LAV}, HIV_{LAI}, HIV_{PH}, HIV-1_{CR215}, HIV-1_{USA}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2_{UC1} and HIV-2_{UC2}). See, e.g., Myers, et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico; Myers, et al., *Human Retroviruses and Aids*, 1990, Los Alamos, New Mexico: Los Alamos National Laboratory.
- Proteins derived from other viruses will also find use in the claimed methods, such as without limitation, proteins from members of the families Picornaviridae (e.g., polioviruses, etc.); Caliciviridae; Togaviridae (e.g., rubella virus, dengue virus, etc.); Flaviviridae; Coronaviridae; Reoviridae; Birnaviridae; Rhabdoviridae (e.g., rabies virus, etc.); Filoviridae; Paramyxoviridae (e.g., mumps virus, measles virus, respiratory syncytial virus, etc.); Orthomyxoviridae (e.g., influenza virus types A, B and C, etc.); Bunyaviridae; Arenaviridae; Retroviridae, e.g., HTLV-I; HTLV-II; HIV-1; HIV-2; simian immunodeficiency virus (SIV) among others. See, e.g. Virology, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991; Virology, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA) for a description of these and other viruses.

Particularly preferred bacterial antigens are derived from organisms that cause diphtheria, tetanus, pertussis, meningitis, and other pathogenic states, including, without limitation, antigens derived from *Corynebacterium diphtheriae*, *Clostridium tetani*, *Bordetella pertussis*, *Neisseria meningitidis*, including serotypes *Meningococcus* A, B, C, Y and W135 (MenA, B, C, Y and W135), *Haemophilus influenza* type B (Hib), and

Helicobacter pylori. Examples of parasitic antigens include those derived from organisms causing malaria, tuberculosis, and Lyme disease.

Furthermore, the methods described herein provide
5 means for treating a variety of malignant cancers. For example, the system of the present invention can be used to enhance both humoral and cell-mediated immune responses to particular proteins specific to a cancer in question, such as an activated oncogene, a fetal antigen,
10 or an activation marker. Such tumor antigens include any of the various MAGEs (melanoma associated antigen E), including MAGE 1, 2, 3, 4, etc. (Boon, T. *Scientific American* (March 1993):82-89); any of the various tyrosinases; MART 1 (melanoma antigen recognized by T
15 cells), mutant ras; mutant p53; p97 melanoma antigen; CEA (carcinoembryonic antigen), among others.

DNA immunization using synthetic expression
cassettes of the present invention has been demonstrated to be efficacious (Examples 8 and 10-12). Animals were
20 immunized with both the synthetic expression cassette and the wild type expression cassette. The results of the immunizations with plasmid-DNAs showed that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression
25 cassettes. Also, the second boost immunization induced a secondary immune response, for example after two to eight weeks. Further, the results of CTL assays showed increased potency of synthetic expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by
30 DNA immunization.

It is readily apparent that the subject invention can be used to mount an immune response to a wide variety of antigens and hence to treat or prevent a large number of diseases.

2.3.1 DELIVERY OF THE SYNTHETIC EXPRESSION CASSETTES OF THE
PRESENT INVENTION

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant
5 methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. The sequences can be analyzed by conventional sequencing techniques. Furthermore, the desired gene can be isolated directly
10 from cells and tissues containing the same, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain, isolate and sequence DNA. Once the sequence is known, the gene
15 of interest can also be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the intended host in which the sequence will
20 be expressed. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge, *Nature* (1981) 292:756; Nambair et al., *Science* (1984) 223:1299; Jay et al., *J. Biol. Chem.* (1984) 259:6311; Stemmer, W.P.C., (1995) *Gene*
25 164:49-53.

Next, the gene sequence encoding the desired antigen can be inserted into a vector containing a synthetic expression cassette of the present invention (e.g., see
30 Example 1 for construction of various exemplary synthetic expression cassette). The antigen is inserted into the synthetic coding sequence such that when the combined sequence is expressed it results in the production of VLPs comprising the polypeptide and/or the antigen of

interest. Insertions can be made within the Gag coding sequence or at either end of the coding sequence (5', amino terminus of the expressed polypeptide; or 3', carboxy terminus of the expressed polypeptide -- e.g., see Example 1) (Wagner, R., et al., *Arch Virol.* 127:117-137, 1992; Wagner, R., et al., *Virology* 200:162-175, 1994; Wu, X., et al., *J. Virol.* 69(6):3389-3398, 1995; Wang, C-T., et al., *Virology* 200:524-534, 1994; Chazal, N., et al., *Virology* 68(1):111-122, 1994; Griffiths, J.C., et al., *J. Virol.* 67(6):3191-3198, 1993; Reicin, A.S., et al., *J. Virol.* 69(2):642-650, 1995).

Up to 50% of the coding sequences of p55Gag can be deleted without affecting the assembly to virus-like particles and expression efficiency (Borsetti, A., et al, *J. Virol.* 72(11):9313-9317, 1998; Gamier, L., et al., *J Virol* 72(6):4667-4677, 1998; Zhang, Y., et al., *J Virol* 72(3):1782-1789, 1998; Wang, C., et al., *J Virol* 72(10):7950-7959, 1998). In one embodiment of the present invention, immunogenicity of the high level expressing synthetic p55GagMod and p55GagProtMod expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted, mutated or truncated regions of p55GagMod sequence. In another embodiment of the present invention, immunogenicity of the high level expressing synthetic Env expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted regions of gp120Mod, gp140Mod or gp160Mod sequences. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length modified Env

sequences fused to heterologous polypeptides, can be higher expression/secretion efficiency and/or higher immunogenicity of the expression product. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length Env, Gag or Tat sequences fused to heterologous polypeptides, can be higher expression/secretion efficiency and/or immunogenicity of the expression product.

When sequences are added to the amino terminal end of Gag (for example, when using the synthetic p55GagMod expression cassette of the present invention), the polynucleotide can contain coding sequences at the 5' end that encode a signal for addition of a myristic moiety to the Gag-containing polypeptide (e.g., sequences that encode Met-Gly).

The ability of Gag-containing polypeptide constructs to form VLPs can be empirically determined following the teachings of the present specification.

HIV polypeptide/antigen synthetic expression cassettes include control elements operably linked to the coding sequence, which allow for the expression of the gene *in vivo* in the subject species. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter, the mouse mammary tumor virus LTR promoter, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop

codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook et al., *supra*, as well as a bovine growth hormone terminator sequence.

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence.

Furthermore, plasmids can be constructed which include a chimeric antigen-coding gene sequences, encoding, e.g., multiple antigens/epitopes of interest, for example derived from a single or from more than one viral isolate.

Typically the antigen coding sequences precede or follow the synthetic coding sequences and the chimeric transcription unit will have a single open reading frame encoding both the antigen of interest and the synthetic Gag coding sequences. Alternatively, multi-cistronic cassettes (e.g., bi-cistronic cassettes) can be constructed allowing expression of multiple antigens from a single mRNA using the EMCV IRES, or the like. Lastly, antigens can be encoded on separate transcripts from independent promoters on a single plasmid or other vector.

Once complete, the constructs are used for nucleic acid immunization or the like using standard gene

delivery protocols. Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859, 5,589,466. Genes can be delivered either directly to the vertebrate subject or, alternatively, delivered ex vivo, to cells derived from the subject and the cells reimplanted in the subject.

A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery systems. Selected sequences can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either in vivo or ex vivo. A number of retroviral systems have been described (U.S. Patent No. 5,219,740; Miller and Rosman, *BioTechniques* (1989) 7:980-990; Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa et al., *Virology* (1991) 180:849-852; Burns et al., *Proc. Natl. Acad. Sci. USA* (1993) 20:8033-8037; and Boris-Lawrie and Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109).

A number of adenovirus vectors have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol.* (1986) 57:267-274; Bett et al., *J. Virol.* (1993) 67:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) 5:717-729; Seth et al., *J. Virol.* (1994) 68:933-940; Barr et al., *Gene Therapy* (1994) 1:51-58; Berkner, K.L. *BioTechniques* (1988) 6:616-629; and Rich et al., *Human Gene Therapy* (1993) 4:461-476).

Additionally, various adeno-associated virus (AAV) vector systems have been developed for gene delivery.

AAV vectors can be readily constructed using techniques well known in the art. See, e.g., U.S. Patent Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 (published 23 January 1992) and WO 93/03769 (published 4 March 1993); Lebkowski et al., *Molec. Cell. Biol.* (1988) 8:3988-3996; Vincent et al., *Vaccines 90* (1990) (Cold Spring Harbor Laboratory Press); Carter, B.J. *Current Opinion in Biotechnology* (1992) 3:533-539; Muzyczka, N. *Current Topics in Microbiol. and Immunol.* (1992) 158:97-129; Kotin, R.M. *Human Gene Therapy* (1994) 5:793-801; Shelling and Smith, *Gene Therapy* (1994) 1:165-169; and Zhou et al., *J. Exp. Med.* (1994) 179:1867-1875.

Another vector system useful for delivering the polynucleotides of the present invention is the enterically administered recombinant poxvirus vaccines described by Small, Jr., P.A., et al. (U.S. Patent No. 5,676,950, issued October 14, 1997).

Additional viral vectors which will find use for delivering the nucleic acid molecules encoding the antigens of interest include those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the genes can be constructed as follows. The DNA encoding the particular synthetic Gag/antigen coding sequence is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the sequences of interest into the viral genome. The resulting TK recombinant can be selected by culturing the

cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the
5 genes. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an avipox vector is particularly desirable in human and other mammalian species since members of the avipox
10 genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the
15 production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for
20 gene delivery.

Members of the Alphavirus genus, such as, but not limited to, vectors derived from the Sindbis, Semliki Forest, and Venezuelan Equine Encephalitis viruses, will
25 also find use as viral vectors for delivering the polynucleotides of the present invention (for example, a synthetic Gag- or Env-polypeptide encoding expression cassette as described in Example 14 below). For a description of Sindbis-virus derived vectors useful for
30 the practice of the instant methods, see, Dubensky et al., *J. Virol.* (1996) 70:508-519; and International Publication Nos. WO 95/07995 and WO 96/17072; as well as, Dubensky, Jr., T.W., et al., U.S. Patent No. 5,843,723,

issued December 1, 1998, and Dubensky, Jr., T.W., U.S. Patent No. 5,789,245, issued August 4, 1998.

A vaccinia based infection/transfection system can be conveniently used to provide for inducible, transient
5 expression of the coding sequences of interest (for example, a synthetic Gag/HCV-core expression cassette) in a host cell. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase
10 displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus
15 recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See,
20 e.g., Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA* (1990) 87:6743-6747; Fuerst et al., *Proc. Natl. Acad. Sci. USA* (1986) 83:8122-8126.

As an alternative approach to infection with vaccinia or avipox virus recombinants, or to the delivery
25 of genes using other viral vectors, an amplification system can be used that will lead to high level expression following introduction into host cells. Specifically, a T7 RNA polymerase promoter preceding the coding region for T7 RNA polymerase can be engineered.
30 Translation of RNA derived from this template will generate T7 RNA polymerase which in turn will transcribe more template. Concomitantly, there will be a cDNA whose expression is under the control of the T7 promoter. Thus, some of the T7 RNA polymerase generated from

translation of the amplification template RNA will lead to transcription of the desired gene. Because some T7 RNA polymerase is required to initiate the amplification, T7 RNA polymerase can be introduced into cells along with the template(s) to prime the transcription reaction. The polymerase can be introduced as a protein or on a plasmid encoding the RNA polymerase. For a further discussion of T7 systems and their use for transforming cells, see, e.g., International Publication No. WO 94/26911; Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130; Deng and Wolff, *Gene* (1994) 143:245-249; Gao et al., *Biochem. Biophys. Res. Commun.* (1994) 200:1201-1206; Gao and Huang, *Nuc. Acids Res.* (1993) 21:2867-2872; Chen et al., *Nuc. Acids Res.* (1994) 22:2114-2120; and U.S. Patent No. 5,135,855.

The synthetic expression cassette of interest can also be delivered without a viral vector. For example, the synthetic expression cassette can be packaged as DNA or RNA in liposomes prior to delivery to the subject or to cells derived therefrom. Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed DNA to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight, *Biochim. Biophys. Acta.* (1991) 1097:1-17; Straubinger et al., in *Methods of Enzymology* (1983), Vol. 101, pp. 512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations, with cationic liposomes particularly preferred. Cationic liposomes have been shown to mediate intracellular

delivery of plasmid DNA (Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416); mRNA (Malone et al., *Proc. Natl. Acad. Sci. USA* (1989) 86:6077-6081); and purified transcription factors (Debs et al., *J. Biol. Chem.* (1990) 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY.

(See, also, Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416). Other commercially available lipids include (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g., Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; PCT Publication No. WO 90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as, from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilamellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See, e.g., Straubinger et al., in *METHODS OF*

- IMMUNOLOGY (1983), Vol. 101, pp. 512-527; Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; Papahadjopoulos et al., *Biochim. Biophys. Acta* (1975) 394:483; Wilson et al., *Cell* (1979) 17:77; Deamer and
- 5 Bangham, *Biochim. Biophys. Acta* (1976) 443:629; Ostro et al., *Biochem. Biophys. Res. Commun.* (1977) 76:836; Fraley et al., *Proc. Natl. Acad. Sci. USA* (1979) 76:3348; Enoch and Strittmatter, *Proc. Natl. Acad. Sci. USA* (1979) 76:145; Fraley et al., *J. Biol. Chem.* (1980) 255:10431;
- 10 Szoka and Papahadjopoulos, *Proc. Natl. Acad. Sci. USA* (1978) 75:145; and Schaefer-Ridder et al., *Science* (1982) 215:166.

- The DNA and/or protein antigen(s) can also be delivered in cochleate lipid compositions similar to
- 15 those described by Papahadjopoulos et al., *Biochem. Biophys. Acta*. (1975) 394:483-491. See, also, U.S. Patent Nos. 4,663,161 and 4,871,488.

- The synthetic expression cassette of interest (e.g., any of the synthetic expression cassettes described in
- 20 Example 1) may also be encapsulated, adsorbed to, or associated with, particulate carriers. Such carriers present multiple copies of a selected antigen to the immune system and promote migration, trapping and retention of antigens in local lymph nodes. The
- 25 particles can be taken up by profession antigen presenting cells such as macrophages and dendritic cells, and/or can enhance antigen presentation through other mechanisms such as stimulation of cytokine release. Examples of particulate carriers include those derived
- 30 from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP,

et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993.

Furthermore, other particulate systems and polymers can be used for the *in vivo* or *ex vivo* delivery of the gene of interest. For example, polymers such as polylysine, polyarginine, polyornithine, spermine, spermidine, as well as conjugates of these molecules, are useful for transferring a nucleic acid of interest. Similarly, DEAE dextran-mediated transfection, calcium phosphate precipitation or precipitation using other insoluble inorganic salts, such as strontium phosphate, aluminum silicates including bentonite and kaolin, chromic oxide, magnesium silicate, talc, and the like, will find use with the present methods. See, e.g., Felgner, P.L., *Advanced Drug Delivery Reviews* (1990) 5:163-187, for a review of delivery systems useful for gene transfer. Peptoids (Zuckerman, R.N., et al., U.S. Patent No. 5,831,005, issued November 3, 1998) may also be used for delivery of a construct of the present invention.

Additionally, biolistic delivery systems employing particulate carriers such as gold and tungsten, are especially useful for delivering synthetic expression cassettes of the present invention. The particles are coated with the synthetic expression cassette(s) to be delivered and accelerated to high velocity, generally under a reduced atmosphere, using a gun powder discharge from a "gene gun." For a description of such techniques, and apparatuses useful therefore, see, e.g., U.S. Patent Nos. 4,945,050; 5,036,006; 5,100,792; 5,179,022; 5,371,015; and 5,478,744. Also, needle-less injection systems can be used (Davis, H.L., et al, *Vaccine* 12:1503-1509, 1994; Bioject, Inc., Portland, OR).

Recombinant vectors carrying a synthetic expression cassette of the present invention are formulated into compositions for delivery to the vertebrate subject. These compositions may either be prophylactic (to prevent
5 infection) or therapeutic (to treat disease after infection). The compositions will comprise a "therapeutically effective amount" of the gene of interest such that an amount of the antigen can be produced *in vivo* so that an immune response is generated
10 in the individual to which it is administered. The exact amount necessary will vary depending on the subject being treated; the age and general condition of the subject to be treated; the capacity of the subject's immune system to synthesize antibodies; the degree of protection
15 desired; the severity of the condition being treated; the particular antigen selected and its mode of administration, among other factors. An appropriate effective amount can be readily determined by one of skill in the art. Thus, a "therapeutically effective
20 amount" will fall in a relatively broad range that can be determined through routine trials.

The compositions will generally include one or more "pharmaceutically acceptable excipients or vehicles" such as water, saline, glycerol, polyethyleneglycol,
25 hyaluronic acid, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, surfactants and the like, may be present in such vehicles. Certain facilitators of immunogenicity or of nucleic acid uptake and/or
30 expression can also be included in the compositions or coadministered, such as, but not limited to, bupivacaine, cardiotoxin and sucrose.

Once formulated, the compositions of the invention can be administered directly to the subject (e.g., as

described above) or, alternatively, delivered *ex vivo*, to cells derived from the subject, using methods such as those described above. For example, methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and can include, e.g., dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, lipofectamine and LT-1 mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) (with or without the corresponding antigen) in liposomes, and direct microinjection of the DNA into nuclei.

Direct delivery of synthetic expression cassette compositions *in vivo* will generally be accomplished with or without viral vectors, as described above, by injection using either a conventional syringe, needless devices such as Bioject® or a gene gun, such as the Accell® gene delivery system (PowderJect Technologies, Inc., Oxford, England). The constructs can be delivered (e.g., injected) either subcutaneously, epidermally, intradermally, intramuscularly, intravenous, intramucosally (such as nasally, rectally and vaginally), intraperitoneally or orally. Delivery of DNA into cells of the epidermis is particularly preferred as this mode of administration provides access to skin-associated lymphoid cells and provides for a transient presence of DNA in the recipient. Other modes of administration include oral ingestion and pulmonary administration, suppositories, needle-less injection, transcutaneous and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule.

2.3.2 Ex vivo DELIVERY OF THE SYNTHETIC EXPRESSION

CASSETTES OF THE PRESENT INVENTION

In one embodiment, T cells, and related cell types (including but not limited to antigen presenting cells, such as, macrophage, monocytes, lymphoid cells, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof), can be used for ex vivo delivery of the synthetic expression cassettes of the present invention. T cells can be isolated from peripheral blood lymphocytes (PBLs) by a variety of procedures known to those skilled in the art. For example, T cell populations can be "enriched" from a population of PBLs through the removal of accessory and B cells. In particular, T cell enrichment can be accomplished by the elimination of non-T cells using anti-MHC class II monoclonal antibodies. Similarly, other antibodies can be used to deplete specific populations of non-T cells. For example, anti-Ig antibody molecules can be used to deplete B cells and anti-MacI antibody molecules can be used to deplete macrophages.

T cells can be further fractionated into a number of different subpopulations by techniques known to those skilled in the art. Two major subpopulations can be isolated based on their differential expression of the cell surface markers CD4 and CD8. For example, following the enrichment of T cells as described above, CD4⁺ cells can be enriched using antibodies specific for CD4 (see Coligan et al., *supra*). The antibodies may be coupled to a solid support such as magnetic beads. Conversely, CD8⁺ cells can be enriched through the use of antibodies specific for CD4 (to remove CD4⁺ cells), or can be isolated by the use of CD8 antibodies coupled to a solid support. CD4

lymphocytes from HIV-1 infected patients can be expanded ex vivo, before or after transduction as described by Wilson et. al. (1995) *J. Infect. Dis.* 172:88.

5 Following purification of T cells, a variety of methods of genetic modification known to those skilled in the art can be performed using non-viral or viral-based gene transfer vectors constructed as described herein. For example, one such approach involves transduction of
10 the purified T cell population with vector-containing supernatant of cultures derived from vector producing cells. A second approach involves co-cultivation of an irradiated monolayer of vector-producing cells with the purified T cells. A third approach involves a similar
15 co-cultivation approach; however, the purified T cells are pre-stimulated with various cytokines and cultured 48 hours prior to the co-cultivation with the irradiated vector producing cells. Pre-stimulation prior to such transduction increases effective gene transfer (Nolta et
20 al. (1992) *Exp. Hematol.* 20:1065). Stimulation of these cultures to proliferate also provides increased cell populations for re-infusion into the patient. Subsequent to co-cultivation, T cells are collected from the vector producing cell monolayer, expanded, and frozen in liquid
25 nitrogen.

Gene transfer vectors, containing one or more synthetic expression cassette of the present invention (associated with appropriate control elements for delivery to the isolated T cells) can be assembled using
30 known methods.

Selectable markers can also be used in the construction of gene transfer vectors. For example, a marker can be used which imparts to a mammalian cell transduced with the gene transfer vector resistance to a

cytotoxic agent. The cytotoxic agent can be, but is not limited to, neomycin, aminoglycoside, tetracycline, chloramphenicol, sulfonamide, actinomycin, netropsin, distamycin A, anthracycline, or pyrazinamide. For
5 example, neomycin phosphotransferase II imparts resistance to the neomycin analogue geneticin (G418).

The T cells can also be maintained in a medium containing at least one type of growth factor prior to being selected. A variety of growth factors are known in
10 the art which sustain the growth of a particular cell type. Examples of such growth factors are cytokine mitogens such as rIL-2, IL-10, IL-12, and IL-15, which promote growth and activation of lymphocytes. Certain types of cells are stimulated by other growth factors
15 such as hormones, including human chorionic gonadotropin (hCG) and human growth hormone. The selection of an appropriate growth factor for a particular cell population is readily accomplished by one of skill in the art.

20 For example, white blood cells such as differentiated progenitor and stem cells are stimulated by a variety of growth factors. More particularly, IL-3, IL-4, IL-5, IL-6, IL-9, GM-CSF, M-CSF, and G-CSF, produced by activated T_H and activated macrophages,
25 stimulate myeloid stem cells, which then differentiate into pluripotent stem cells, granulocyte-monocyte progenitors, eosinophil progenitors, basophil progenitors, megakaryocytes, and erythroid progenitors. Differentiation is modulated by growth factors such as
30 GM-CSF, IL-3, IL-6, IL-11, and EPO.

Pluripotent stem cells then differentiate into lymphoid stem cells, bone marrow stromal cells, T cell progenitors, B cell progenitors, thymocytes, T_H Cells, T_C cells, and B cells. This differentiation is modulated by

growth factors such as IL-3, IL-4, IL-6, IL-7, GM-CSF, M-CSF, G-CSF, IL-2, and IL-5.

Granulocyte-monocyte progenitors differentiate to monocytes, macrophages, and neutrophils. Such differentiation is modulated by the growth factors GM-CSF, M-CSF, and IL-8. Eosinophil progenitors differentiate into eosinophils. This process is modulated by GM-CSF and IL-5.

The differentiation of basophil progenitors into mast cells and basophils is modulated by GM-CSF, IL-4, and IL-9. Megakaryocytes produce platelets in response to GM-CSF, EPO, and IL-6. Erythroid progenitor cells differentiate into red blood cells in response to EPO.

Thus, during activation by the CD3-binding agent, T cells can also be contacted with a mitogen, for example a cytokine such as IL-2. In particularly preferred embodiments, the IL-2 is added to the population of T cells at a concentration of about 50 to 100 $\mu\text{g/ml}$. Activation with the CD3-binding agent can be carried out for 2 to 4 days.

Once suitably activated, the T cells are genetically modified by contacting the same with a suitable gene transfer vector under conditions that allow for transfection of the vectors into the T cells. Genetic modification is carried out when the cell density of the T cell population is between about 0.1×10^6 and 5×10^6 , preferably between about 0.5×10^6 and 2×10^6 . A number of suitable viral and nonviral-based gene transfer vectors have been described for use herein.

After transduction, transduced cells are selected away from non-transduced cells using known techniques. For example, if the gene transfer vector used in the transduction includes a selectable marker which confers resistance to a cytotoxic agent, the cells can be

contacted with the appropriate cytotoxic agent, whereby non-transduced cells can be negatively selected away from the transduced cells. If the selectable marker is a cell surface marker, the cells can be contacted with a binding agent specific for the particular cell surface marker, whereby the transduced cells can be positively selected away from the population. The selection step can also entail fluorescence-activated cell sorting (FACS) techniques, such as where FACS is used to select cells from the population containing a particular surface marker, or the selection step can entail the use of magnetically responsive particles as retrievable supports for target cell capture and/or background removal.

More particularly, positive selection of the transduced cells can be performed using a FACS cell sorter (e.g. a FACSVantage™ Cell Sorter, Becton Dickinson Immunocytometry Systems, San Jose, CA) to sort and collect transduced cells expressing a selectable cell surface marker. Following transduction, the cells are stained with fluorescent-labeled antibody molecules directed against the particular cell surface marker. The amount of bound antibody on each cell can be measured by passing droplets containing the cells through the cell sorter. By imparting an electromagnetic charge to droplets containing the stained cells, the transduced cells can be separated from other cells. The positively selected cells are then harvested in sterile collection vessels. These cell sorting procedures are described in detail, for example, in the FACSVantage™ Training Manual, with particular reference to sections 3-11 to 3-28 and 10-1 to 10-17.

Positive selection of the transduced cells can also be performed using magnetic separation of cells based on expression of a particular cell surface marker. In such

separation techniques, cells to be positively selected are first contacted with specific binding agent (e.g., an antibody or reagent that interacts specifically with the cell surface marker). The cells are then contacted with
5 retrievable particles (e.g., magnetically responsive particles) which are coupled with a reagent that binds the specific binding agent (that has bound to the positive cells). The cell-binding agent-particle complex can then be physically separated from non-labeled cells,
10 for example using a magnetic field. When using magnetically responsive particles, the labeled cells can be retained in a container using a magnetic field while the negative cells are removed. These and similar separation procedures are known to those of ordinary
15 skill in the art.

Expression of the vector in the selected transduced cells can be assessed by a number of assays known to those skilled in the art. For example, Western blot or Northern analysis can be employed depending on the nature
20 of the inserted nucleotide sequence of interest. Once expression has been established and the transformed T cells have been tested for the presence of the selected synthetic expression cassette, they are ready for infusion into a patient via the peripheral blood stream.

25 The invention includes a kit for genetic modification of an *ex vivo* population of primary mammalian cells. The kit typically contains a gene transfer vector coding for at least one selectable marker and at least one synthetic expression cassette contained
30 in one or more containers, ancillary reagents or hardware, and instructions for use of the kit.

EXPERIMENTAL

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

Example 1Generation of Synthetic Gag and Env Expression Cassettes

15 A. Modification of HIV-1 Gag, Gag-protease, Gag-reverse transcriptase and Gag-polymerase Nucleic Acid Coding Sequences

The Gag (SEQ ID NO:1), Gag-protease (SEQ ID NO:2), Gag-polymerase (SEQ ID NO:3), and Gag-reverse transcriptase (SEQ ID NO:77) coding sequences were selected from the HIV-1SF2 strain (Sanchez-Pescador, R., et al., *Science* 227(4686): 484-492, 1985; Luciw, P.A., et al. U.S. Patent No. 5,156,949, issued October 20, 1992; Luciw, P.A., et al., U.S. Patent No. 5,688,688, November 18, 1997). These sequences were manipulated to maximize expression of their gene products.

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a high AU content in the RNA and in a decreased translation ability and instability of the

mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Gag-encoding sequences were modified to be comparable to codon usage found in highly expressed human genes.

5 Figure 11 presents a comparison of the percent A-T content for the cDNAs of stable versus unstable RNAs (comparison window size = 50). Human IFN γ mRNA is known to (i) be unstable, (ii) have a short half-life, and (iii) have a high A-U content. Human GAPDH
10 (glyceraldehyde-3-phosphate dehydrogenase) mRNA is known to (i) be a stable RNA, and (i) have a low A-U content. In Figure 11, the percent A-T content of these two sequences are compared to the percent A-T content of native HIV-1SF2 Gag cDNA and to the synthetic Gag cDNA
15 sequence of the present invention. The top two panels of the figure show the percent A-T content over the length of the sequences for IFN γ and native Gag. The bottom two panels of the figure show the percent A-T content over the length of the sequences for GAPDH and the synthetic
20 Gag. Experiments performed in support of the present invention showed that the synthetic Gag sequences were capable of higher level of protein production (see the Examples) than the native Gag sequences. The data in Figure 11 suggest that one reason for this increased
25 production may be increased stability of the mRNA corresponding to the synthetic Gag coding sequences versus the mRNA corresponding to the native Gag coding sequences.

 Second, there are inhibitory (or instability)
30 elements (INS) located within the coding sequences of the Gag and Gag-protease coding sequences (Schneider R, et al., *J Virol.* 71(7):4892-4903, 1997). RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating

effects of the INS. To overcome the requirement for post-transcriptional activating mechanisms of RRE and Rev, and to enhance independent expression of the Gag polypeptide, the INS were inactivated by introducing
5 multiple point mutations that did not alter the reading frame of the encoded proteins. Figure 1 shows the original SF2 Gag sequence, the location of the INS sequences, and the modifications made to the INS sequences to reduce their effects.

10 For the Gag-protease sequence (wild type, SEQ ID NO:2; synthetic, SEQ ID NOS:5, 78 and 79), the changes in codon usage were restricted to the regions up to the -1 frameshift and starting again at the end of the Gag reading frame (Figure 2; the region indicated in lower
15 case letters in Figure 2 is the unmodified region). Further, inhibitory (or instability) elements (INS) located within the coding sequences of the Gag-protease polypeptide coding sequence were altered as well (indicated in Figure 2). The synthetic coding sequences
20 were assembled by the Midland Certified Reagent Company (Midland, Texas).

Modification of the Gag-polymerase sequences (wild type, SEQ ID NO:3; synthetic, SEQ ID NO:6) and Gag-reverse transcriptase sequences (SEQ ID NOS:80 through
25 84) include similar modifications as described for Gag-protease in order to preserve the frameshift region. Locations of the inactivation sites and changes to the sequence to alter the inactivation sites are presented in Figure 12 for the native HIV-1_{sp2} Gag-polymerase sequence.

30 In one embodiment of the invention, the full length polymerase coding region of the Gag-polymerase sequence is included with the synthetic Gag sequences in order to increase the number of epitopes for virus-like particles expressed by the synthetic, optimized Gag expression

- cassette. Because synthetic HIV-1 Gag-polymerase expresses the potentially deleterious functional enzymes reverse transcriptase (RT) and integrase (INT) (in addition to the structural proteins and protease), it is important to inactivate RT and INT functions. Several in-frame deletions in the RT and INT reading frame can be made to achieve catalytic nonfunctional enzymes with respect to their RT and INT activity. {Jay. A. Levy (Editor) (1995) *The Retroviridae*, Plenum Press, New York. ISBN 0-306-45033X. Pages 215-20; Grimison, B. and Laurence, J. (1995), *Journal Of Acquired Immune Deficiency Syndromes and Human Retrovirology* 9(1):58-68; Wakefield, J. K., et al., (1992) *Journal Of Virology* 66(11):6806-6812; Esnouf, R., et al., (1995) *Nature Structural Biology* 2(4):303-308; Maignan, S., et al., (1998) *Journal Of Molecular Biology* 282(2):359-368; Katz, R. A. and Skalka, A. M. (1994) *Annual Review Of Biochemistry* 73 (1994); Jacobo-Molina, A., et al., (1993) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 90(13):6320-6324; Hickman, A. B., et al., (1994) *Journal Of Biological Chemistry* 269(46):29279-29287; Goldgur, Y., et al., (1998) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 95(16):9150-9154; Goette, M., et al., (1998) *Journal Of Biological Chemistry* 273(17):10139-10146; Gorton, J. L., et al., (1998) *Journal of Virology* 72(6):5046-5055; Engelman, A., et al., (1997) *Journal Of Virology* 71(5):3507-3514; Dyda, F., et al., *Science* 266(5193):1981-1986; Davies, J. F., et al., (1991) *Science* 252(5002):88-95; Bujacz, G., et al., (1996) *Febs Letters* 398(2-3):175-178; Beard, W. A., et al., (1996) *Journal Of Biological Chemistry* 271(21):12213-12220; Kohlstaedt, L. A., et al., (1992)

Science 256(5065):1783-1790; Krug, M. S. and Berger, S. L. (1991) *Biochemistry* 30(44):10614-10623; Mazumder, A., et al., (1996) *Molecular Pharmacology* 49(4):621-628; Palaniappan, C., et al., (1997) *Journal Of Biological*
5 *Chemistry* 272(17):11157-11164; Rodgers, D. W., et al., (1995) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 92(4):1222-1226; Sheng, N. and Dennis, D. (1993) *Biochemistry* 32(18):4938-4942; Spence, R. A., et al., (1995) *Science* 267(5200):988-993.}

10 Furthermore selected B- and/or T-cell epitopes can be added to the Gag-polymerase constructs within the deletions of the RT- and INT-coding sequence to replace and augment any epitopes deleted by the functional modifications of RT and INT. Alternately, selected B-
15 and T-cell epitopes (including CTL epitopes) from RT and INT can be included in a minimal VLP formed by expression of the synthetic Gag or synthetic GagProt cassette, described above. (For descriptions of known HIV B- and T-cell epitopes see, HIV Molecular Immunology Database CTL
20 Search Interface; Los Alamos Sequence Compendia, 1987-1997; Internet address: <http://hiv-web.lanl.gov/immunology/index.html>.)

The resulting modified coding sequences are presented as a synthetic Gag expression cassette (SEQ ID
25 NO:4), a synthetic Gag-protease expression cassette (SEQ ID NOs:5, 78 and 79), and a synthetic Gag-polymerase expression cassette (SEQ ID NO:6). Synthetic expression cassettes containing codon modifications in the reverse transcriptase region are shown in SEQ ID NOs:80 through
30 84. An alignment of selected sequences is presented in Figure 7. A common region (Gag-common; SEQ ID NO:9) extends from position 1 to position 1262.

The synthetic DNA fragments for Gag and Gag-protease were cloned into the following expression vectors:

pCMVKm2, for transient expression assays and DNA immunization studies, the pCMVKm2 vector was derived from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) and comprises a kanamycin selectable marker, a ColE1 origin of replication, a CMV promoter enhancer and Intron A, followed by an insertion site for the synthetic sequences described below followed by a polyadenylation signal derived from bovine growth hormone -- the pCMVKm2 vector differs from the pCMV-link vector only in that a polylinker site was inserted into pCMVKm2 to generate pCMV-link (Figure 14, polylinker at positions 1646 to 1697); pESN2dhfr (Figure 13A) and pCMVPLEdhfr (also known as pCMVIII as shown in Figure 13B), for expression in Chinese Hamster Ovary (CHO) cells; and, pAcc13, a shuttle vector for use in the Baculovirus expression system (pAcc13, was derived from pAcc12 which was described by Munemitsu S., et al., *Mol Cell Biol.* 10(11):5977-5982, 1990).

A restriction map for vector pCMV-link is presented in Figure 14. In the figure, the CMV promoter (CMV IE ENH/PRO), bovine growth hormone terminator (BGH pA), kanamycin selectable marker (kan), and a ColE1 origin of replication (ColE1 ori) are indicated. A polycloning site is also indicated in the figure following the CMV promoter sequences.

A restriction map for vector pESN2dhfr is presented in Figure 13A. In the figure, the CMV promoter (pCMV, hCMVIE), bovine growth hormone terminator (BGH pA), SV40 origin of replication (SV40ori), neomycin selectable marker (Neo), SV40 polyA (SV40pA), Adenovirus 2 late promoter (Ad2VLP), and the murine dhfr gene (mu dhfr) are indicated. A polycloning site is also indicated in the figure following the CMV promoter sequences.

Briefly, construction of pCMVPLEdhfr (pCMVIII) was as follows. To construct a DHFR cassette, the EMCV IRES (internal ribosome entry site) leader was PCR-amplified from pCite-4a+ (Novagen, Inc., Milwaukee, WI) and
 5 inserted into pET-23d (Novagen, Inc., Milwaukee, WI) as an *Xba*-*Nco* fragment to give pET-EMCV. The *dhfr* gene was PCR-amplified from pESN2dhfr to give a product with a Gly-Gly-Gly-Ser spacer in place of the translation stop codon and inserted as an *Nco*-*Bam*H1 fragment to give pET-
 10 E-DHFR. Next, the attenuated *neo* gene was PCR amplified from a pSV2Neo (Clontech, Palo Alto, CA) derivative and inserted into the unique *Bam*H1 site of pET-E-DHFR to give pET-E-DHFR/Neo_(n2). Then, the bovine growth hormone terminator from pCDNA3 (Invitrogen, Inc., Carlsbad, CA)
 15 was inserted downstream of the *neo* gene to give pET-E-DHFR/Neo_(n2)BGht. The EMCV-*dhfr*/*neo* selectable marker cassette fragment was prepared by cleavage of pET-E-DHFR/Neo_(n2)BGht. The CMV enhancer/promoter plus Intron A was transferred from pCMV6a (Chapman et al., *Nuc. Acids*
 20 *Res.* (1991) 19:3979-3986) as a *Hind*III-*Sal*I fragment into pUC19 (New England Biolabs, Inc., Beverly, MA). The vector backbone of pUC19 was deleted from the *Nde*I to the *Sap*I sites. The above described DHFR cassette was added to the construct such that the EMCV IRES followed the CMV
 25 promoter to produce the final construct. The vector also contained an amp^r gene and an SV40 origin of replication.

Selected pCMVKm2 vectors containing the synthetic expression cassettes have been designated as follows: pCMVKm2.GagMod.SF2, pCMVKm2.GagprotMod.SF2, and
 30 pCMVKm2.GagpolMod.SF2, pCMVKm2.GagprotMod.SF2.GP1 (SEQ ID NO:78) and pCMVKm2.GagprotMod.SF2.GP2 (SEQ ID NO:79). Other exemplary Gag-encoding expressing cassettes are shown in the Figures and as Sequence Listings.

E. Modification of HIV-1 Gag/Hepatitis C Core Chimeric Protein Nucleic Acid Coding Sequences Generation of Synthetic Expression Cassettes

To facilitate the ligation of the Gag and HCV core coding sequences, PCR amplification was employed. The synthetic p55Gag expression cassette was used as a PCR template with the following primers: GAG5 (SEQ ID NO:11) and P55-SAL3 (SEQ ID NO:12). The PCR amplification was conducted at 55°C for 25 cycles using Stratagene's Pfu polymerase. The resulting PCR product was rendered free of nucleotides and primers using the Promega PCR clean-up kit and then subjected to EcoRI and SalI digestions. For HCV core coding sequences, the following primers were used with an HCV template (Houghton, M., et al., U.S. Patent No. 5,714,596, issued February 3, 1998; Houghton, M., et al., U.S. Patent No. 5,712,088, issued January 27, 1998; Houghton, M., et al., U.S. Patent No. 5,683,864, issued November 4, 1997; Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J., et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No. 5,670,152, issued September 23, 1997): CORESAL 5 (SEQ ID NO:13) and 173CORE (SEQ ID NO:14) using the conditions outlined above. The purified product was digested with SalI and BamHI restriction enzymes. The digested Gag and HCV core PCR products were ligated into the pCMVKm2 vector digested with EcoRI and BamHI. Ligation of the PCR products at the SalI site resulted in a direct fusion of the final amino acid of p55Gag to the second amino acid of HCV core, serine. Amino acid 173 of core is a serine and is followed immediately by a TAG termination codon. The sequence of the fusion clone was confirmed. The pCMVKm2 vector containing the synthetic expression

cassette was designated as pCMVKm2.GagModHCVcore.

The EcoRI-BamHI fragment of p55Gag-core 173 was also cloned into EcoRI-BamHI-digested pAcC13 for baculovirus expression. Western blots confirmed expression and
5 sucrose gradient sedimentation along with electron microscopy confirmed particle formation. To generate the above clone but containing the synthetic Gag sequences (instead of wild-type), the following steps were performed: pCMVKm2-modified p55Gag was used as template
10 for PCR amplification with MS65 (SEQ ID NO:15) and MS66(SEQ ID NO:16) primers. The region amplified corresponds to the BspHI and SalI sites at the C-terminus of synthetic Gag sequence. The amplification product was digested with BspHI and SalI and ligated to SalI/BamHI
15 digested pCMV-link along with the Sal/BspHI fragment from pCMV-Km-p55modGag, representing the amino terminal end of modified Gag, and the SalI/BamHI fragment from pCMV-p55Gag-core173. Thereafter, a T4-blunted-SalI partial/BamHI fragment was ligated into pAcC4-SmaI/BamHI
20 to generate pAcC4-p55GagMod-core173 (containing the synthetic sequence presented as SEQ ID NO:7).

C. Defining of the Major Homology Region (MHR) of HIV-1 p55Gag

25 The Major Homology Region (MHR) of HIV-1 p55 (Gag) is located in the p24-CA sequence of Gag. It is a conserved stretch of 20 amino acids (SEQ ID NO:19). The position in the wild type HIV-1_{sf2} Gag protein is from aa 286-305 and spans a region from nucleotides 856-915 in
30 the native HIV-1_{sf2} Gag DNA-sequence. The position in the synthetic Gag protein is from aa 288-307 and spans a region from nucleotides 862-921 for the synthetic Gag DNA-sequence. The nucleotide sequence for the MHR in the synthetic

GagMod.SF2 is presented as SEQ ID NO:20. Mutations or deletions in the amino acid sequence of the MHR can severely impair particle production (Borsetti, A., et al., *J. Virol.* 72(11):9313-9317, 1998; Mammano, F., et al., *J Virol* 68(8):4927-4936, 1994).

Percent identity to the MHR nucleotide sequence can be determined, for example, using the MacDNAsis program (Hitachi Software Engineering America Limited, South San Francisco, CA), Higgins algorithm, with the following exemplary parameters: gap penalty = 5, no. of top diagonals = 5, fixed gap penalty = 5, K-tuple = 2, window size = 5, and floating gap penalty = 10.

D. Generation of Synthetic Env Expression Cassettes

Env coding sequences of the present invention include, but are not limited to, polynucleotide sequences encoding the following HIV-encoded polypeptides: gp160, gp140, and gp120 (see, e.g., U.S. Patent No. 5,792,459 for a description of the HIV-1_{SF2} ("SF2") Env polypeptide). The relationships between these polypeptides is shown schematically in Figure 15 (in the figure: the polypeptides are indicated as lines, the amino and carboxy termini are indicated on the gp160 line; the open circle represents the oligomerization domain; the open square represents a transmembrane spanning domain (TM); and "c" represents the location of a cleavage site, in gp140.mut the "X" indicates that the cleavage site has been mutated such that it no longer functions as a cleavage site). The polypeptide gp160 includes the coding sequences for gp120 and gp41. The polypeptide gp41 is comprised of several domains including an oligomerization domain (OD) and a transmembrane spanning domain (TM). In the native envelope, the oligomerization domain is required for the

non-covalent association of three gp41 polypeptides to form a trimeric structure: through non-covalent interactions with the gp41 trimer (and itself), the gp120 polypeptides are also organized in a trimeric structure.

5 A cleavage site (or cleavage sites) exists approximately between the polypeptide sequences for gp120 and the polypeptide sequences corresponding to gp41. This cleavage site(s) can be mutated to prevent cleavage at the site. The resulting gp140 polypeptide corresponds to

10 a truncated form of gp160 where the transmembrane spanning domain of gp41 has been deleted. This gp140 polypeptide can exist in both monomeric and oligomeric (i.e. trimeric) forms by virtue of the presence of the oligomerization domain in the gp41 moiety. In the

15 situation where the cleavage site has been mutated to prevent cleavage and the transmembrane portion of gp41 has been deleted the resulting polypeptide product is designated "mutated" gp140 (e.g., gp140.mut). As will be apparent to those in the field, the cleavage site can be

20 mutated in a variety of ways. The native amino acid sequence in the SF162 cleavage sites is: APTKAKRRVVQREKR (SEQ ID NO:21), where KAKRR (SEQ ID NO:22) is termed the "second" site and REKR (SEQ ID NO:23) is the "first site". Exemplary mutations include the following

25 constructs: gp140.mut7.modSF162 which encodes the amino acid sequence APTKA**ISSVVQSEKS** (SEQ ID NO:24) in the cleavage site region; gp140.mut8.modSF162 which encodes the amino acid sequence APT**IAISSVVQSEKS** (SEQ ID NO:25) in the cleavage site region and gp140mut.modSF162 which

30 encodes the amino acid sequence APTKAKRRVVQREKS (SEQ ID NO:26). Mutations are denoted in bold. The native amino acid sequence in the US4 cleavage sites is: APTQAKRRVVQREKR (SEQ ID NO:27), where QAKRR (SEQ ID NO:28) is termed the "second" site and REKR (SEQ ID

NO:23) is the "first site". Exemplary mutations include the following construct: gp140.mut.modUS4 which encodes the amino acid sequence APTQAKRRVVQREKS (SEQ ID NO:29) in the cleavage site region. Mutations are denoted in bold.

5

E. Modification of HIV-1 Env (Envelope) Nucleic Acid Coding Sequences

In one embodiment of the present invention, wild-type Env coding sequences were selected from the HIV-1^{SF162} ("SF162") strain (Cheng-Mayer (1989) *PNAS USA* 86:8575-8579). These SF162 sequences were as follows: gp120, SEQ ID NO:30 (Fig. 16); gp140, SEQ ID NO:31 (Fig. 17); and gp160, SEQ ID NO:32 (Fig. 18).

In another embodiment of the present invention, wild-type Env coding sequences were selected from the HIV-US4 strain (Mascola, et al. (1994) *J. Infect. Dis.* 169:48-54). These US4 sequences were as follows: gp120, SEQ ID NO:51 (Fig. 38); gp140, SEQ ID NO:52 (Fig. 39); and gp160, SEQ ID NO:53 (Fig. 40).

These Env coding sequences were manipulated to maximize expression of their gene products.

First, the wild-type coding region was modified in one or more of the following ways. In one embodiment, sequences encoding hypervariable regions of Env, particularly V1 and/or V2 were deleted. In other embodiments, mutations were introduced into sequences encoding the cleavage site in Env to abrogate the enzymatic cleavage of oligomeric gp140 into gp120 monomers. (See, e.g., Earl et al. (1990) *PNAS USA* 87:648-652; Earl et al. (1991) *J. Virol.* 65:31-41). In yet other embodiments, hypervariable region(s) were deleted, N-glycosylation sites were removed and/or cleavage sites mutated.

Second, the HIV-1 codon usage pattern was modified

so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T in the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Env coding sequences were modified to be comparable to codon usage found in highly expressed human genes.

Figures 22A-22H present comparisons of the percent A-T content for the cDNAs of stable versus unstable RNAs (comparison window size = 50). Human IFN γ mRNA is known to (i) be unstable, (ii) have a short half-life, and (iii) have a high A-U content. Human GAPDH (glyceraldehyde-3-phosphate dehydrogenase) mRNA is known to (i) be a stable RNA, and (i) have a low A-U content. In Figures 22A-H, the percent A-T content of these two sequences are compared to the percent A-T content of (1) native HIV-1 US4 Env gp160 cDNA, a synthetic US4 Env gp160 cDNA sequence (i.e., having modified codons) of the present invention; and (2) native HIV-1 SF162 Env gp160 cDNA, a synthetic SF162 Env gp160 cDNA sequence (i.e., having modified codons) of the present invention. Figures 22A-H show the percent A-T content over the length of the sequences for IFN γ (Figures 22C and 22G); native gp160 Env US4 and SF162 (Figures 22A and 22E, respectively); GAPDH (Figures 22D and 22H); and the synthetic gp160 Env for US4 and SF162 (Figures 22B and 22F). Experiments performed in support of the present invention showed that the synthetic Env sequences were capable of higher level of protein production (see the Examples) than the native Env sequences. The data in Figures 22A-H suggest that one reason for this increased

production is increased stability of the mRNA corresponding to the synthetic Env coding sequences versus the mRNA corresponding to the native Env coding sequences.

- 5 To create the synthetic coding sequences of the present invention the gene cassettes were designed to comprise the entire coding sequence of interest. Synthetic gene cassettes were constructed by oligonucleotide synthesis and PCR amplification to
10 generate gene fragments. Primers were chosen to provide convenient restriction sites for subcloning. The resulting fragments were then ligated to create the entire desired sequence which was then cloned into an appropriate vector. The final synthetic sequences were
15 (i) screened by restriction endonuclease digestion and analysis, (ii) subjected to DNA sequencing in order to confirm that the desired sequence had been obtained and (iii) the identity and integrity of the expressed protein confirmed by SDS-PAGE and Western blotting (See,
20 Examples. The synthetic coding sequences were assembled at Chiron Corp. or by the Midland Certified Reagent Company (Midland, Texas).

- Exemplary modified coding sequences are presented as synthetic Env expression cassettes in Table 1A and 1B.
25 The following expression cassettes (i) have unique, terminal *EcoRI* and *XbaI* cloning sites; (ii) include Kozak sequences to promote optimal translation; (iii) tPA signal sequences (to direct the ENV polypeptide to the cell membrane, see, e.g., Chapman et al., *infra*); (iv)
30 open reading frames optimized for expression in mammalian cells; and (v) a translational stop signal codon.

Table 1A: Exemplary Synthetic Env Expression
Cassettes (SF162)

	Expression Cassette	Seq Id	Further Information
5	gp120 SF162	30	wild-type; Figure 16
	gp140 SF162	31	wild-type; Figure 17
	gp160 SF162	32	wild-type; Figure 18
	gp120.modSF162	33	none; Figure 19
	gp120.modSF162.delV2	34	deleted V2 loop; Figure 20
10	gp120.modSF162.delV1/V2	35	deleted V1 and V2; Figure 21
	gp140.modSF162	36	none; Figure 23
	gp140.modSF162.delV2	37	deleted V2 loop; Figure 24
	gp140.modSF162.delV1/V2	38	deleted V1 and V2; Figure 25
	gp140.mut.modSF162	39	mutated cleavage site; Fig. 26
15	gp140.mut.modSF162.delV2	40	deleted V2; mutated cleavage site; Figure 27
	gp140.mut.modSF162.delV1/V2	41	deleted V1 & V2; mutated cleavage site; Figure 28
	gp140.mut7.modSF162	42	mutated cleavage site; Fig. 29
	gp140.mut7.modSF162.delV2	43	mutated cleavage site; deleted V2; Figure 30
20	gp140.mut7.modSF162.delV1/V2	44	mutated cleavage site; deleted V1 and V2; Figure 31
	gp140.mut8.modSF162	45	mutated cleavage site; Fig. 32
	gp140.mut8.modSF162.delV2	46	mutated cleavage site; deleted V2; Figure 33
25	gp140.mut8.modSF162.delV1/V2	47	mutated cleavage site; deleted V1 and V2; Figure 34
	gp160.modSF162	48	none; Figure 35
	gp160.modSF162.delV2	49	deleted V2 loop; Figure 36
	gp160.modSF162.delV1/V2	50	deleted V1 & V2; Figure 37

Table 1B:
Exemplary Synthetic Env Expression Cassettes(US4)

	Expression Cassette	Seq Id	Further Information
5	gp120 US4	51	wild-type; Figure 38
	gp140 US4	52	wild-type; Figure 39
	gp160 US4	53	wild-type; Figure 40
	gp120.modUS4	54	none; Figure 41
	gp120.modUS4.del 128-194	55	deletion in V1 and V2 regions; Figure 42
10	gp140.modUS4	56	none; Figure 43
	gp140.mut.modUS4	57	mutated cleavage site; Figure 44
	gp140TM.modUS4	58	native transmembrane region; Figure 45
	gp140.modUS4.delV1/V2	59	deleted V1 and V2; Figure 46
	gp140.modUS4.delV2	60	deleted V1; Figure 47
15	gp140.mut.modUS4.delV1/V2	61	mutated cleavage site; deleted V1 and V2; Figure 48
	gp140.modUS4.del 128-194	62	deletion in V1 and V2 regions; Figure 49
	gp140.mut.modUS4.del 128-194	63	mutated cleavage site; deletion in V1 and V2 regions; Figure 50
	gp160.modUS4	64	none; Figure 51
	gp160.modUS4.delV1	65	deleted V1; Figure 52
20	gp160.modUS4.delV2	66	deleted V2; Figure 53
	gp160.modUS4.delV1/V2	67	deleted V1 and V2; Figure 54
	gp160.modUS4del 128-194	68	deletion in V1 and V2 regions; Figure 55

Alignments of the sequences presented in the above
tables are presented in Figures 66A and 66B.

A common region (Env-common) extends from nucleotide position 1186 to nucleotide position 1329 (SEQ ID NO:69,

Fig. 56) relative to the wild-type US4 sequence and from nucleotide position 1117 to position 1260 (SEQ ID NO:79, Fig. 57) relative to the wild-type SF162 sequence. The synthetic sequences of the present invention

5 corresponding to these regions are presented, as SEQ ID NO:71 (Figure 58) for the synthetic Env US4 common region and as SEQ ID NO:72 (Figure 59) for the synthetic Env SF162 common region.

Percent identity to this sequence can be determined,

10 for example, using the Smith-Waterman search algorithm (Time Logic, Incline Village, NV), with the following exemplary parameters: weight matrix = nuc4x4hb; gap opening penalty = 20, gap extension penalty = 5, reporting threshold = 1; alignment threshold = 20.

15 Various forms of the different embodiments of the present invention (e.g., constructs) may be combined.

F. Cloning Synthetic Env Expression Cassettes of the Present Invention.

20 The synthetic DNA fragments encoding the Env polypeptides were typically cloned into the eucaryotic expression vectors described above for Gag, for example, pCMVKm2/pCMVlink (Figure 4), pCMV6a, pESN2dhfr (Figure 13A), pCMVIII (Figure 13B; alternately designated as the

25 pCMV-PL-E-dhfr/neo vector).

Exemplary designations for pCMVlink vectors containing synthetic expression cassettes of the present invention are as follows: pCMVlink.gp140.modSF162; pCMVlink.gp140.-modSF162.delV2;

30 pCMVlink.gp140.mut.modSF162; pCMVlink.gp140.mut.modSF162.delV2; pCMVKm2.gp140modUS4; pCMVKm2.gp140.modUS4.delV2; pCMVKm2.gp140.mut.modUS4; and, pCMVKm2.gp140.mut.modUS4.delV1/V2.

G. Generation of Synthetic Tat Expression Cassettes

Tat coding sequences have also been modified according to the teachings of the present specification. The wild type nucleotide sequence encoding tat from variant SF162 is presented in Figure 76 (SEQ ID NO:85). The corresponding wild-type amino acid sequence is presented in Figure 77 (SEQ ID NO:86). Figure 81 (SEQ ID NO:89) shows the nucleotide sequence encoding the amino terminal of the tat protein and the codon encoding cystein-22 is underlined. Other exemplary constructs encoding synthetic tat polypeptides are shown in Figures 78 and 79 (SEQ ID NOs:87 and 88). In one embodiment (SEQ ID NO:88), the cystein residue at position 22 is replaced by a glycine. Caputo et al. (1996) *Gene Therapy* 3:235 have shown that this mutation affects the trans activation domain of Tat.

Various forms of the different embodiments of the invention, described herein, may be combined.

H. Deposit of Vectors

Selected exemplary constructs shown below and described herein are deposited at Chiron Corporation, Emeryville, CA, 94662-8097, and were sent to the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209 on December 27, 1999.

	Plasmid Name	Chiron Deposit #	Date Sent to ATCC
	pcMVgpp160.modUS4	5094	27 Dec 99
	pcMVgpp160delI.modUS4	5095	27 Dec 99
	pcMVgpp160del2.modUS4	5096	27 Dec 99
5	pcMVgpp160del-2.modUS4	5097	27 Dec 99
	pcMVgpp160del128-194.mod.US4	5098	27 Dec 99
	pcMVgpp140mut.modUS4del128-194	5100	27 Dec 99
	pcMVgpp140.mut.mod.US	5101	27 Dec 99
	pcMVgpp160.modSF162	5125	27 Dec 99
10	pcMVgpp160.modSF162.delV2	5126	27 Dec 99
	pcMVgpp160.modSF162.delV1V2	5127	27 Dec 99
	pcMVgpp140.mut.modSF162delV2	5128	27 Dec 99
	pcMVgpp140.mut7.modSF162	5129	27 Dec 99
	pcMVgpp140.mut7.modSF162delV2	5130	27 Dec 99
15	pcMVgpp140.mut8.modSF162	5131	27 Dec 99
	pcMVgpp140.mut8.modSF162delV2	5132	27 Dec 99
	pcMVgpp140.mut8.modSF162delV1V2	5133	27 Dec 99
	pcMVkm2.Gagprot.Mod.SF2.GP1	5150	27 Dec 99
	pcMVkm2.Gagprot.Mod.SF2.GP2	5151	27 Dec 99
20			

Example 2

Expression Assays for the

Synthetic Gag, Env and Tat Coding Sequences

25 A. Gag and Gag-Protease Coding Sequences

The HIV-1SF2 wild-type Gag (SEQ ID NO:1) and Gag-protease (SEQ ID NO:2) sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Gag (SEQ ID NO:4) and
 30 Gag-protease (SEQ ID NOs:5, 78 or 79)) sequences were cloned.

Expression efficiencies for various vectors carrying the HIV-1SF2 wild-type and synthetic Gag sequences were evaluated as follows. Cells from several mammalian cell lines (293, RD, COS-7, and CHO; all obtained from the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209) were transfected with 2 μ g of DNA in transfection reagent LT1 (PanVera Corporation, 545 Science Dr., Madison, WI). The cells were incubated for 5 hours in reduced serum medium (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium was then replaced with normal medium as follows: 293 cells, IMDM, 10% fetal calf serum, 2% glutamine (BioWhittaker, Walkersville, MD); RD and COS-7 cells, D-MEM, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD); and CHO cells, Ham's F-12, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The cells were incubated for either 48 or 60 hours. Supernatants were harvested and filtered through 0.45 μ m syringe filters and, optionally, stored at -20°C.

Supernatants were evaluated using the Coulter p24-assay (Coulter Corporation, Hialeah, FL, US), using 96-well plates coated with a murine monoclonal antibody directed against HIV core antigen. The HIV-1 p24 antigen binds to the coated wells. Biotinylated antibodies against HIV recognize the bound p24 antigen. Conjugated streptavidin-horseradish peroxidase reacts with the biotin. Color develops from the reaction of peroxidase with TMB substrate. The reaction is terminated by addition of 4N H₂SO₄. The intensity of the color is directly proportional to the amount of HIV p24 antigen in a sample.

The results of these expression assays are presented in Tables 2A and 2B. Tables 2A and 2B shows data

obtained using the synthetic Gag-protease expression cassette of SEQ ID NO:5. Similar results were obtained using the Gag-protease expression cassettes of SEQ ID NOs:78 and 79.

5

Table 2: in vitro gag and gagprot p24 expression

5 TABLE 2a. Increased in vitro expression from modified vs. native *gag* plasmids in supernatants and lysates from transiently transfected cells

experiment	native (nat) ^a modified (mod) ^b	supernatant (sup) lysate (lys)	cell line	hours post transfection	total ng p24 (fold increase)
1	nat	sup	293	48	3.4
	mod	sup	293	48	1260 (371)
	nat	sup	293	60	3.2
	mod	sup	293	60	2222 (694)
2	nat	sup	293	60	1.8
	mod	sup	293	60	1740 (966)
3	nat	sup	293	60	1.8
	mod	sup	293	60	580 (322)
4	nat	lys	293	60	1.5
	mod	lys	293	60	85 (57)
1	nat	sup	RD	48	5.6
	mod	sup	RD	48	66 (12)
	nat	sup	RD	60	7.8
	mod	sup	RD	60	70.2 (9)
2	nat	lys	RD	60	1.9
	mod	lys	RD	60	7.8 (4)
1	nat	sup	COS-7	48	0.4
	mod	sup	COS-7	48	33.4 (84)
2	nat	sup	COS-7	48	0.4
	mod	sup	COS-7	48	10 (25)
	nat	lys	COS-7	48	3
	mod	lys	COS-7	48	14 (5)

^a pCMVLink.Gag.SF2.PRE

^b pCMVXm2.GagMod.SF2

5 TABLE 2b. *In vitro* expression from modified gag and gagprotease plasmids in supernatants and lysates from transiently transfected cells

plasmid	supernatant (sup) lysate (lys)	cell line	hours post transfection	total ng p24 ^d
Gag ^a	sup	293	60	760
GagProt(GP1) ^b	sup	293	60	380
GagProt(GP2) ^c	sup	293	60	320
Gag	lys	293	60	78
GagProt(GP1)	lys	293	60	1250
GagProt(GP2)	lys	293	60	400
Gag	sup	COS-7	72	40
GagProt(GP1)	sup	COS-7	72	150
GagProt(GP2)	sup	COS-7	72	290
Gag	lys	COS-7	72	60
GagProt(GP1)	lys	COS-7	72	63
GagProt(GP2)	lys	COS-7	72	58

^a pCMVKm2.GagMod.SF2

^b pCMVKm2.GagProtMod.SF2 (GP1) gagprotease with codon optimization and inactivation of INS in protease

^c pCMVKm2.GagProtMod.SF2 (GP2) gagprotease with only inactivation of INS in protease

^d Shown are representative results from 3 independent experiments for each cell line tested.

The data showed that the synthetic Gag and Gag-protease expression cassettes provided dramatic increases in production of their protein products, relative to the native (HIV-1SF2 wild-type) sequences, when expressed in a variety of cell lines.

B. Env Coding Sequences

The HIV-SF162 ("SF162") wild-type Env (SEQ ID NO:1-3) and HIV-US4 ("US4") wild-type Env (SEQ ID NO:22-24) sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Env sequences were cloned.

Expression efficiencies for various vectors carrying the SF162 and US4 wild-type and synthetic Env sequences were evaluated essentially as described above for Gag except that cell lysates were prepared in 40 μ l lysis buffer (1.0 % NP40, 0.1 M Tris pH 7.5) and frozen at -20°C and capture ELISAs were performed as follows.

For Capture ELISAs, 250 ng of an ammonium sulfate IgG cut of goat polyclonal antibody to gp120SF2/env2-3 was used to coat each well of a 96-well plate (Corning, Corning, NY). Serial dilutions of gp120/SF2 protein (MID 167) were used to set the quantitation curve from which expression of US4 or SF162 gp120 proteins from transfection supernatant and lysates were calculated. Samples were screened undiluted and, optionally, by serial 2-fold dilutions. A human polyclonal antibody to HIV-1 gp120/SF2 was used to detect bound gp120 envelope protein, followed by horse-radish peroxidase (HRP)-labeled goat anti-human IgG conjugates. TMB (Pierce, Rockford, IL) was used as the substrate and the reaction is terminated by addition of 4N H₂SO₄. The reaction was quantified by measuring the optical density (OD) at 450 nm. The intensity of the color is directly

proportional to the amount of HIV gp120 antigen in a sample. Purified SF2 gp120 protein was diluted and used as a standard.

- 5 The results of the transient expression assays are presented in Tables 3 and 4. Table 3 depicts transient expression in 293 cells transfected with a pCMVKm2 vector carrying the Env cassette of interest. Table 4 depicts transient expression in RD cells transfected with a pCMVKm2 vector carrying the Env cassette of interest.

5

Table 3

Native (N) Synthetic(S)	Cell Line	Total sup (ng)	Sup. fold increase (S v. N)	Total cell lysate (ng)	Cell lysate fold increase (S v. N)	Total (ng)	Total fold increase (S v. N)
N-SP120.US4	RD	87		<1		88	
S-SP120.modUS4	RD	690	8	2	5	693	8
N-SP140.US4	RD	526		0		526	
S-SP140.modUS4	RD	1305	2	1	2	1306	2
S-SP140mut.modUS4	RD	35	N/A	25	N/A	60	N/A
S-SP140TM.modUS4	RD	0	N/A	5	N/A	5	N/A
N-SP160.US4	RD	0		8		8	
S-SP160.modUS4	RD	0	0	30	4	30	4

Table 4

CHO Cell Lines Expression Level of US4 Envelope Constructs			
Constructs	CHO Clone #	MTX Level	Expression Level* (ng/ml)
5 gp120.modUS4	1	3.2 μ M	250-450
	2	1.6 μ M	350-450
	3	200nM	230-580
	4	200nM	300-500
gp140.modUS4	1	1 μ M	155-300
	2	1 μ M	100-260
	3	1 μ M	200-430
gp140.mut. modUS4	1	1 μ M	110-270
	2	1 μ M	100-235
	3	1 μ M	100-220
10 gp140.modUS4 .delV1/V2	1	50nM	313-587**
	2	50nM	237-667**
	3	50nM	492-527**
gp140.mut. modUS4.delV1 /V2	1	50nM	46-328**
	2	50nM	82-318**
	3	50nM	204-385**

15 *All samples measured at T-75 flask stage unless otherwise indicated

**at 24 well and 6 well plate stages

***in a three liter bioreactor perfusion culture this clone yielded approximately 2-5 μ g/ml.

The data showed that the synthetic Env and expression cassettes provided a significant increase in production of their protein products, relative to the native (HIV-1SF162 or US4 wild-type) sequences, when
5 expressed in a variety of cell lines.

C. CHO Cell line Env expression data

Chinese hamster ovary (CHO) cells were transfected with plasmid DNA encoding the synthetic HIV-1 gp120 or
10 gp140 proteins (e.g., pESN2dhfr or pCMVIII vector backbone) using Mirus TransIT-LT1 polyamine transfection reagent (Pan Vera) according to the manufacturers instructions and incubated for 96 hours. After 96 hours, media was changed to selective media (F12 special with
15 250 µg/ml G418) and cells were split 1:5 and incubated for an additional 48 hours. Media was changed every 5-7 days until colonies started forming at which time the colonies were picked, plated into 96 well plates and screened by gp120 Capture ELISA. Positive clones were
20 expanded in 24 well plates and screened several times for Env protein production by Capture ELISA, as described above. After reaching confluency in 24 well plates, positive clones were expanded to T25 flasks (Corning, Corning, NY). These were screened several times after
25 confluency and positive clones were expanded to T75 flasks.

Positive T75 clones were frozen in LN2 and the highest expressing clones amplified with 0-5 µM methotrexate (MTX) at several concentrations and plated in
30 100mm culture dishes. Plates were screened for colony formation and all positive clones were again expanded as described above. Clones were expanded and amplified and screened at each step by gp120 capture ELISA. Positive clones were frozen at each methotrexate level. Highest

producing clones were grown in perfusion bioreactors (3L, 100L) for expansion and adaptation to low serum suspension culture conditions for scale-up to larger bioreactors.

- 5 Tables 5 and 6 show Capture ELISA data from CHO cells transfected with pCMVIII vector carrying a cassette encoding synthetic HIV-US4 and SF162 Env polypeptides (e.g., mutated cleavage sites, modified codon usage and/or deleted hypervariable regions). Thus, stably
- 10 transfected CHO cell lines which express Env polypeptides (e.g., gp120, gp140-monomeric, and gp140-oligomeric) have been produced.

Table 5

CHO Cell Lines Expression Level of US4 Envelope Constructs			
Constructs	CHO Clone #	MTX Level	Expression Level* (ng/ml)
5 gp120.modUS4	1	3.2 μ M	250-450
	2	1.6 μ M	350-450
	3	200nM	230-580***
	4	200nM	300-500
gp140.modUS4	1	1 μ M	155-300
	2	1 μ M	100-260
	3	1 μ M	200-430
gp140.mut. modUS4	1	1 μ M	110-270
	2	1 μ M	100-235
	3	1 μ M	100-220
10 gp140.modUS4 .delV1/V2	1	50nM	313-587**
	2	50nM	237-667**
	3	50nM	492-527**
gp140.mut. modUS4.delV1 /V2	1	50nM	46-328**
	2	50nM	82-318**
	3	50nM	204-385**

15 *All samples measured at T-75 flask stage unless otherwise indicated

**at 24 well and 6 well plate stages

***in a three liter bioreactor perfusion culture this clone yielded approximately 2-5 μ g/ml.

Table 6

CHO Cell Lines Expression Level of SF162 Envelope Constructs			
Constructs	CHO Clone #	MTX Level	Expression Level' (ng/ml)
gp120.modSF162	1	0	755-2705
	2	0	928-1538
	3	0	538-1609
gp140.modSF162	1	20 nM	180-350
gp140.mut. modSF162	1	20 nM	164-451
	2	20 nM	188-487
	3	20 nM	233-804
gp120.modSF162 .delV2	1	800nM	528-1560
	2	800nM	487-1878
	3	800nM	589-1212
gp140.modSF162 .delV2	1	800nM	300-600
	2	800nM	200-400
	3	800nM	200-500
gp140.mut. modSF162.delV2	1	800nM	300-700
	2	400nM	1161
	3	800nM	400-600
	4	400nM	1600-2176

15 'All samples measured at T-75 flask stage unless otherwise indicated

The results presented above demonstrate the ability of the constructs of the present invention to provide expression of Env polypeptides in CHO cells. Production of polypeptides using CHO cells provides (i) correct glycosylation patterns and protein conformation (as determined by binding to panel of MABs); (ii) correct binding to CD4 receptor molecules; (iii) absence of non-

mammalian cell contaminants (e.g., insect viruses and/or cells); and (iv) ease of purification.

D. Tat Coding Sequences

5 The HIV-SF162 ("SF162") wild-type Tat (SEQ ID NO:85) sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Tat sequences were cloned (SEQ ID NOs:87, 88 and 89).

10 Expression efficiencies for various vectors carrying the SF162 wild-type and synthetic Tat sequences are evaluated essentially as described above for Gag and Env using capture ELISAs with the appropriate anti-tat antibodies and/or CHO cell assays. Expression of the polypeptides encoded by the synthetic cassettes is
15 improved relative to wild type.

Example 3

Western Blot Analysis of Expression

A. Gag and Gag-Protease Coding Sequences

20 Human 293 cells were transfected as described in Example 2 with pCMV6a-based vectors containing native or synthetic Gag expression cassettes. Cells were cultivated for 60 hours post-transfection. Supernatants were prepared as described. Cell lysates were prepared
25 as follows. The cells were washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO) in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes. SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego,
30 CA) were loaded with 20 μ l of supernatant or 12.5 μ l of cell lysate. A protein standard was also loaded (5 μ l, broad size range standard; BioRad Laboratories, Hercules, CA). Electrophoresis was carried out and the proteins were transferred using a BioRad Transfer Chamber (BioRad

Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer (Millipore), where the transfer was performed at 100 volts for 90 minutes. The
5 membranes were exposed to HIV-1-positive human patient serum and immunostained using o-phenylenediamine dihydrochloride (OPD; Sigma).

The results of the immunoblotting analysis showed that cells containing the synthetic Gag expression
10 cassette produced the expected p55 protein at higher per-cell concentrations than cells containing the native expression cassette. The Gag p55 protein was seen in both cell lysates and supernatants. The levels of production were significantly higher in cell supernatants
15 for cells transfected with the synthetic Gag expression cassette of the present invention. Experiments performed in support of the present invention suggest that cells containing the synthetic Gag-prot expression cassette produced the expected Gag-prot protein at comparably
20 higher per-cell concentrations than cells containing the native expression cassette.

In addition, supernatants from the transfected 293 cells were fractionated on sucrose gradients. Aliquots of the supernatant were transferred to Polyclear™ ultra-
25 centrifuge tubes (Beckman Instruments, Columbia, MD), under-laid with a solution of 20% (wt/wt) sucrose, and subjected to 2 hours centrifugation at 28,000 rpm in a Beckman SW28 rotor. The resulting pellet was suspended in PBS and layered onto a 20-60% (wt/wt) sucrose gradient
30 and subjected to 2 hours centrifugation at 40,000 rpm in a Beckman SW41ti rotor.

The gradient was then fractionated into approximately 10 x 1 ml aliquots (starting at the top, 20%-end, of the gradient). Samples were taken from

fractions 1-9 and were electrophoresed on 8-16% SDS polyacrylamide gels. Fraction number 4 (the peak fraction) corresponds to the expected density of Gag protein VLPs. The supernatants from 293/synthetic Gag cells gave much stronger p55 bands than supernatants from 293/native Gag cells, and, as expected, the highest concentration of p55 in either supernatant was found in fraction 4.

These results demonstrate that the synthetic Gag expression cassette provides superior production of both p55 protein and VLPs, relative to the native Gag coding sequences.

B. Env Coding Sequences

Human 293 cells were transfected as described in Example 2 with pCMVKm2-based; pCMVlink-based; p-CMVII-based or pESN2-based vectors containing native or synthetic Env expression cassettes. Cells were cultivated for 48 or 60 hours post-transfection. Cell lysates and supernatants were prepared as described (Example 2). Briefly, the cells were washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO)] in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes. SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego, CA) were loaded with 20 μ l of supernatant or 12.5 μ l of cell lysate. A protein molecular weight standard and an HIV SF2 gp120 positive control protein (5 μ l, broad size range standard; BioRad Laboratories, Hercules, CA) were also loaded. Electrophoresis was carried out and the proteins were transferred using a BioRad Transfer Chamber (BioRad Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer

(Millipore), where the transfer was performed at 100 volts for 90 minutes. The membranes were then reacted against polyclonal goat anti-gp120SF2/env2-3 anti-sera, followed by incubation with swine anti-goat IgG-
5 peroxidase (POD) (Sigma, St. Louis, MO). Bands indicative of binding were visualized by adding DAB with hydrogen peroxide which deposits a brown precipitate on the membranes.

The results of the immunoblotting analysis showed
10 that cells containing the synthetic Env expression cassette produced the expected Env gp proteins of the predicted molecular weights as determined by mobilities in SDS-polyacrylamide gels at higher per-cell concentrations than cells containing the native
15 expression cassette. The Env proteins were seen in both cell lysates and supernatants. The levels of production were significantly higher in cell supernatants for cells transfected with the synthetic Env expression cassette of the present invention.

20 C. Tat Coding Sequences

Human 293 cells are transfected as described in Example 2 with various vectors containing native or synthetic Tat expression cassettes. Cells are cultivated
25 and isolated proteins analyzed as described above. Immunoblotting analysis shows that cells containing the synthetic Tat expression cassette produced the expected Tat proteins of the predicted molecular weights as determined by mobilities in SDS-polyacrylamide gels at
30 higher per-cell concentrations than cells containing the native expression cassette.

Example 4

Purification of Env polypeptides

A. Purification of Oligomeric gp140

Purification of oligomeric gp140 (o-gp140 US4) was
5 conducted essentially as shown in Figure 60. For the
experiments described herein, o-gp140 refers to
oligomeric gp140 in either native or modified (e.g.,
optimized expression sequences, deleted, mutated,
truncated, etc.) form. Briefly, concentrated (30-50X)
10 supernatants obtained from CHO cell cultures were loaded
onto an anion exchange (DEAE) column which removed DNA
and other serum proteins. The eluted material was loaded
onto a ceramic hydroxyapatite column (CHAP) which bound
serum proteins but not HIV Env proteins. The flow-
15 through from the DEAE and CHAP columns was loaded onto a
Protein A column as a precautionary step to remove any
remaining serum immunoglobulins. The Env proteins in the
flow-through were then captured using the lectin
gluconanthus naavalis (GNA, Vector Labs, Burlingame, CA).
20 GNA has high affinity for mannose rich carbohydrates such
as Env. The Env proteins were then eluted with GNA
substrate. To remove other highly glycosylated proteins,
a cation exchange column (SP) was used to purify
gp140/gp120. In a final step, which separates gp120 from
25 o-gp140, a gel filtration column was used to separate
oligomers from monomers. Sizing and chromatography
analysis of the final product revealed that this strategy
lead to the successful isolation of oligomeric gp140.

B. Purification of gp120

Purification of gp120 was conducted essentially as
previously described for other Env proteins. Briefly,
concentrated supernatants obtained from CHO cell cultures
were loaded onto an anion exchange (DEAE) column which

removed DNA and other serum proteins. The eluted material was loaded onto a ceramic hydroxyapatite column (CHAP) which bound serum proteins but not HIV Env proteins. The flow-through from the CHAP column was
5 loaded a cation exchange column (SP) where the flow-through was discarded and the bound fraction eluted with salt. The eluted fraction(s) were loaded onto a Suprose 12/Superdex 200 Tandem column (Pharmacia-Upjohn, Uppsala, Sweden) from which purified gp120 was obtained. Sizing
10 and chromatography analysis of the final product revealed that this strategy successfully purified gp120 proteins.

Example 5

Analysis of Purified Env Polypeptides

15 A. Analysis of o-gp140

It is well documented that HIV Env protein binds to CD4 only in its correct conformation. Accordingly, the ability of o-gp140 US4 polypeptides, produced and purified as described above, to bind CD4 cells was
20 tested. O-gp140 US4 was incubated for 15 minutes with FITC-labeled CD4 at room temperature and loaded onto a Biosil 250 (BioRad) size exclusion column using Waters HPLC. CD4-FITC has the longest retention time (2.67 minutes), followed by CD4-FITC-gp120 (2.167 min). The
25 shortest retention time (1.9 min) was observed for CD4-FITC-o-gp140 US4 indicating that, as expected, o-gp140 US4 binds to CD4 forming a large complex which reduces retention time on the column. Thus, the o-gp140 US4
30 produced and purified as described above is of the correct size and conformation.

In addition, the US4 o-gp140, purified as described above, was also tested for its ability to bind to a variety of monoclonal antibodies with known epitope specificities for the CD4 binding site, the CD4 inducible

site, the V3 loop and oligomer-specific gp41 epitope. O-gp140 bound strongly to these antibodies, indicating that the purified protein retains its structural integrity.

5 B. Analysis of gp120

As described above, CD4-FITC binds gp120, as demonstrated by the decreased retention time on the HPLC column. Thus, US4 gp120 purified by the above method retains its conformational integrity. In addition, the properties of purified gp120 can be tested by examining its integrity and identity on western blots, as well as, by examining protein concentration, pH, conductivity, endotoxin levels, bioburden and the like. US4 gp120, purified as described above, was also tested for its ability to bind to a variety of monoclonal antibodies with known epitope specificities for the CD4 binding site, the CD4 inducible site, the V3 loop and oligomer-specific gp41 epitope. The pattern of mAb binding to gp120 indicated that the purified protein retained its structural integrity, for example, the purified gp120 did not bind the mAb having the oligomer-specific gp41 epitope (as expected).

Example 6

25 Electron Microscopic Evaluation of VLP Production

The cells for electron microscopy were plated at a density of 50-70% confluence, one day before transfection. The cells were transfected with 10 µg of DNA using transfection reagent LT1 (Panvera) and incubated for 5 hours in serum-reduced medium (see Example 2). The medium was then replaced with normal medium (see Example 2) and the cells were incubated for 14 hours (COS-7) or 40 hours (CHO). After incubation the cells were washed twice with PBS and fixed with 2%

glutaraldehyde. Electron microscopy was performed by Prof. T.S. Benedict Yen, Veterans Affairs, Medical Center, San Francisco, CA).

Electron microscopy was carried out using a transmission electron microscope (Zeiss 10c). The cells were pre-stained with osmium and stained with uranium acetate and lead citrate. The magnification was 100,000X.

Figures 3A and 3B show micrographs of CHO cells transfected with pCMVKM2 carrying the synthetic Gag expression cassette (SEQ ID NO:5) or carrying the Gag-prot expression cassette (SEQ ID NO:79). In the figure, free and budding immature virus-like-particles (VLP) of the expected size (100 nm) are seen for the Gag expression cassette (Figure 3A) and both immature and mature VLPs are seen for the Gag-prot expression cassette (Figure 3B). COS-7 cells transfected with the same vector have the same expression pattern. VLP can also be found intracellularly in CHO and COS-7 cells.

Native and synthetic Gag expression cassettes were compared for their associated levels of VLP production when used to transfect human 293 cells. The comparison was performed by density gradient ultracentrifugation of cell supernatants and Western-blot analysis of the gradient fractions. There was a clear improvement in production of VLPs when using the synthetic Gag construct.

Example 7

Expression of Virus-like Particles in the Baculovirus System

A. Expression of Native HIV p55 Gag

To construct the native HIV p55 Gag baculovirus shuttle vector, the prototype SF2 HIV p55 plasmid, pTM1-

Gag (Selby M.J., et al., *J Virol.* 71(10):7827-7831, 1997), was digested with restriction endonucleases NcoI and BamHI to extract a 1.5 Kb fragment that was subsequently subcloned into pAcC4 (*Bio/Technology* 6:47-55, 1988), a derivative of pAc436. Generation of the recombinant baculovirus was achieved by co-transfecting 2 μ g of the HIV p55 Gag pAcC4 shuttle vector with 0.5 μ g of linearized, *Autographa californica* baculovirus (AcNPV) wild-type viral DNA into *Spodoptera frugiperda* (Sf9) cells (Kitts, P.A., Ayres M.D., and Possee R.D., *Nucleic Acids Res.* 18:5667-5672, 1990). The isolation of recombinant virus expressing HIV p55 Gag was performed according to standard techniques (O'Reilly, D.R., L.K. Miller, and V. A. Luckow, *Baculovirus Expression Vector: A Laboratory Manual*, W.H. Freeman and Company, New York, 1992).

Expression of the HIV p55 Gag was achieved using a 500 ml suspension culture of Sf9 cells grown in serum-free medium (Miaorella, B., D. Inlow, A. Shauger, and D. Harano, *Bio/Technology* 6:1506-1510, 1988) that had been infected with the HIV p55 Gag recombinant baculovirus at a multiplicity of infection (MOI) of 10. Forty-eight hours post-infection, the supernatant was separated by centrifugation and filtered through a 0.2 μ m filter. Aliquots of the supernatant were then transferred to Polyclear™ (Beckman Instruments, Palo Alto, CA) ultracentrifuge tubes, underlaid with 20% (wt/wt) sucrose, and subjected to 2 hours centrifugation at 24,000 rpm using a Beckman SW28 rotor.

The resulting pellet was suspended in Tris buffer (20 mM Tris HCl, pH 7.5, 250 mM NaCl, and 2.5 mM ethylenediaminetetraacetic acid [EDTA]), layered onto a 20-60% (wt/wt) sucrose gradient, and subjected to 2 hours centrifugation at 40,000 rpm using a Beckman SW41ti

rotor. The gradient was then fractionated starting at the top (20% sucrose) of the gradient into approximately twelve 0.75 ml aliquots. A sample of each fraction was electrophoresed on 8-16% SDS polyacrylamide gels and the resulting bands were visualized after commassie staining (Figure 4). Additional aliquots were subjected to refractive index analysis.

The results shown in Figure 4 indicated that the p55 Gag virus-like particles banded at a sucrose density of range of 1.15 - 1.19 g/ml with the peak at approximately 1.17 g/ml. The peak fractions were pooled and concentrated by a second 20% sucrose pelleting. The resulting pellet was suspended in 1 ml of Tris buffer (described above). The total protein yield as estimated by Bicinchrominic Acid (BCA) (Pierce Chemical, Rockford, IL) was 1.6 mg.

B. Expression of Synthetic HIV p55 Gag

A baculovirus shuttle vector containing the synthetic p55 Gag sequence was constructed as follows. The synthetic HIV p55 expression cassette (Example 1) was digested with restriction enzyme *SalI* followed by incubation with T4-DNA polymerase. The resulting fragment was isolated (PCR Clean-Up™, Promega, Madison, WI) and then digested with *BamHI* endonuclease. The shuttle vector pAcCl3 (Munemitsu S., et al., *Mol Cell Biol.* 10(11):5977-5982, 1990) was linearized by digestion with *EcoI*, followed by incubation with T4-DNA polymerase, and then isolated (PCR Clean-Up™). The linearized vector was digested with *BamHI*, treated with alkaline phosphatase, and isolated by size fragmentation in an agarose gel. The isolated 1.5 kb fragment was ligated with the prepared pAcCl3 vector. The resulting clone was designated pAcCl3-Modif.p55Gag.

The expression conditions for the synthetic HIV p55 VLPs differed from those of the native p55 Gag as follows: a culture volume of 1 liter used instead of 500 ml; *Trichoplusia ni* (Tn5) (Wickham, T.J., and Nermerow, G.R., *BioTechnology Progress*, 9:25-30, 1993) insect cells were used instead of Sf9 insect cells; and, an MOI of 3 was instead of an MOI of 10. Experiments performed in support of the present invention showed that there was no appreciable difference in expression level between the Sf9 and Tn5 insect cells with the native p55 clone. In terms of MOI, experience with the native p55 clone suggested that an MOI of 10 resulted in higher expression (approximately 2-fold) of VLPs than a lower MOI.

The sucrose pelleting and banding methods used for the synthetic p55 VLPs were similar to those employed for the native p55 VLPs (described above), with the following exceptions: pelleted VLPs were suspended in 4 ml of phosphate buffered saline (PBS) instead of 1.0 ml of the Tris buffer; and four, 20-60% sucrose gradients were used instead of a single gradient. Also, due to the high concentration of banded VLPs, further concentration by pelleting was not required. The peak fractions from all 4 gradients were simply dialyzed against PBS. The approximate density of the banded VLPs ranged from 1.23-1.28 g/ml. A total protein yield as estimated by BCA was 46 mg. Results from the sucrose gradient banding of the synthetic p55 are shown in Figure 5.

A comparison of the total amount of purified HIV p55 Gag from several preparations obtained from the two baculovirus expression cassettes has been summarized in Figure 6. The average yield from the native p55 was 3.16 mg/liter of culture (n=5, standard deviation (sd) ± 1.07 , range = 1.8-4.8 mg/L) whereas the average yield from the

synthetic p55 was more than ten-fold higher at 44.5 mg/liter of culture ($n=2$, $sd=\pm 6.4$).

In addition to a higher total protein yield, the final product from the synthetic p55-expressed Gag consistently contained lower amounts of contaminating baculovirus proteins than the final product from the native p55-expressed Gag. This difference can be seen in the two commassie-stained gels Figures 4 and 5.

10 C. Expression of Native and Synthetic Gag-Core

Expression of the HIV p55 Gag/HCV Core 173 (SEQ ID NO:8) was achieved using a 2.5 liter suspension culture of Sf9 cells grown in serum-free medium (Miaorella, B., D. Inlow, A. Shauger, and D. Harano. 1988 Bio/Technology 6:1506-1510). The cells were infected with an HIV p55 Gag/HCV Core 173 recombinant baculovirus. Forty-eight hours post-infection, the supernatant was separated from the cells by centrifugation and filtered through a 0.2 μ m filter. Aliquots of the supernatant were then transferred to a Polyclear™ (Beckman Instruments, Palo Alto, CA) ultracentrifuge tubes containing 30% (wt/wt) sucrose, and subjected to 2 hours of centrifugation at 24,000 rpm in a Beckman SW28 rotor and ultracentrifuge.

The resulting pellet was suspended in Tris buffer (50 mM Tris-HCl, pH 7.5, 500 mM NaCl) and layered onto a 30-60% (wt/wt) sucrose gradient and subjected to 2 hours centrifugation at 40,000 rpm in a Beckman SW41ti rotor and ultracentrifuge. The gradient was then fractionated starting at the top (30%) of the gradient into approximately 11 x 1.0 ml aliquots. A sample of each fraction was electrophoresed on 8-16% SDS polyacrylamide gels and the resulting bands were visualized after commassie staining.

A subset of aliquots were also subjected to Western blot analysis using monoclonal antibody 76C.5EG (Steimer, K.S., et al., Virology 150:283-290, 1986) which is specific for HIV p24 (a subunit of HIV p55). The peak
5 fractions from the sucrose gradient were pooled and concentrated by a second 20% sucrose pelleting. The resulting pellet was suspended in 1 ml of buffer Tris buffer and the total protein yield as estimated by BCA (Pierce Chemical, Rockford, IL) was ~ 1.0 mg.

10 The results from the SDS PAGE are shown in Figure 8 and the anti- p24 Western blot results are shown in Figure 9. Taken together, these results indicate that the HIV p55 Gag/HCV Core 173 chimeric VLPs banded at a sucrose density similar to that of the HIV p55 Gag VLPs
15 and the visible protein band that migrated at a molecular weight of ~ 72,000 kd was reactive with the HIV p24-specific monoclonal antibody. An additional immunoreactive band at approximately 55,000 kd also appeared to be reactive with the anti-p24 antibody and
20 may be a degradation product.

Although aliquots from the above preparation were not tested for reactivity with an HCV Core-specific antibody (an anti-CD22 rabbit serum), results from a similar preparation are shown in Figure 10 and indicate
25 that the main HCV Core-specific reactivity migrates at an approximate molecular weight of 72,000 kd which is in accordance with the predicted molecular weight of the chimeric protein.

The expression conditions for the synthetic HIV p55
30 Gag/HCV Core 173 (SEQ ID NO:8) VLPs differed from those of the native p55 Gag and are as follows: a culture volume of 1 liter used instead of 2.5 liters, *Trichoplusia ni* (Tn5) (Wickham, T.J., and Nemerow, G.R. 1993 BioTechnology Progress, 9:25-30) insect cells were

used instead of Sf9 insect cells and an MOI of 3 was instead of an MOI of 10. The sucrose pelleting and banding methods used for the synthetic HIV p55 Gag/HCV Core 173 VLPs were similar to those employed for the native HIV p55 Gag/HCV Core 173 VLPs. However, differences included: pelleted VLPs were suspended in 1 ml of phosphate buffered saline (PBS) instead of 1.0 ml of the Tris buffer, and a single 20-60% sucrose gradients was used. A comparison of the total amount of purified HIV p55 Gag/HCV Core 173 from multiple preparations obtained from the two baculovirus expression cassettes showed that there was an increase in expression using the synthetic HIV p55 Gag/HCV Core 173 cassette.

D. Alternative method for the enrichment of HIV p55 Gag VLPs

In addition to purification from the media, p55 (Gag protein) expressed in baculovirus (e.g., using a synthetic expression cassette of the present invention) can also be purified as virus-like particles from the infected insect cells. For example, forty-eight hours post infection, the media and cell pellet are separated by centrifugation and the cell pellet is stored at -70°C until future use. At the time of processing, the cell pellet is suspended in 5 volumes of hypotonic lysis buffer (20 mM Tris-HCl, pH 8.2, 1 mM EGTA; 1 mM MgCl₂, and Complete Protease Inhibitor® (Boehringer Mannheim Corp., Indianapolis, IN)). If needed, the cells are then dounced 8-10 times to complete cell lysis.

The lysate is then centrifuged at approximately 1000-1500 x g for 20 minutes. The supernatant is

decanted into UltraClear™ tubes, underlayered with 20% sucrose (w/w) and centrifuged at 24,000 rpm in SW28 buckets for 2 hours. The resulting pellet is suspended in Tris buffer (20 mM Tris HCl, pH 7.5, 250 mM NaCl, and 2.5 mM ethylene-
5 diamine-tetraacetic acid (EDTA) with 0.1% IGEPAL detergent (Sigma Chemical, St. Louis, MO) and 250 units/ml of benzonase (American International Chemical, Inc., Natick, MA) and
10 incubated at 4°C for at least 30 minutes. The suspension is subsequently layered onto a 20-60% sucrose gradient and spun at 40,000 rpm using an SW41ti rotor for 20-24 hours.

After ultracentrifugation, the sucrose gradient is
15 fractionated and aliquots run on SDS PAGE to identify peak fractions. The peak fractions are dialyzed against PBS and measured for protein content. Negatively stained electron micrographs typically show non-enveloped VLPs
20 somewhat smaller in diameter (80-120 nm) than the budded VLPs. HIV Gag VLPs prepared in this manner are also capable of generating Gag-specific CTL responses in mice.

Example 8

25 In Vivo Immunogenicity of Synthetic Gag Expression

Cassettes

A. Immunization

To evaluate the possibly improved immunogenicity of the synthetic Gag expression cassettes, a mouse study was
30 performed. The plasmid DNA, pCMVCM2 carrying the synthetic Gag expression cassette, was diluted to the following final concentrations in a total injection volume of 100 µl: 20 µg, 2 µg, 0.2 µg, and 0.02 µg. To

overcome possible negative dilution effects of the diluted DNA, the total DNA concentration in each sample was brought up to 20 μ g using the vector (pCMVKM2) alone. As a control, plasmid DNA of the native Gag expression cassette was handled in the same manner. Twelve groups of four Balb/c mice (Charles River, Boston, MA) were intramuscularly immunized (50 μ l per leg, intramuscular injection into the *tibialis anterior*) according to the schedule in Table 7.

Table 7

Group	Gag Expression Cassette	Concentration of Gag plasmid DNA (μ g)	Immunized at time (weeks):
1	Synthetic	20	0 ¹ , 4
2	Synthetic	2	0, 4
3	Synthetic	0.2	0, 4
4	Synthetic	0.02	0, 4
5	Synthetic	20	0
6	Synthetic	2	0
7	Synthetic	0.2	0
8	Synthetic	0.02	0
9	Native	20	0
10	Native	2	0
11	Native	0.2	0
12	Native	0.02	0

¹ = initial immunization at "week 0"

Groups 1-4 were bled at week 0 (before immunization), week 4, week 6, week 8, and week 12. Groups 5-12 were bled at week 0 (before immunization) and at week 4.

E. Humoral Immune Response

The humoral immune response was checked with an anti-HIV Gag antibody ELISAs (enzyme-linked immunosorbent assays) of the mice sera 0 and 4 weeks post immunization (groups 5-12) and, in addition, 6 and 8 weeks post immunization, respectively, 2 and 4 weeks post second immunization (groups 1-4).

The antibody titers of the sera were determined by anti-Gag antibody ELISA. Briefly, sera from immunized mice were screened for antibodies directed against the HIV p55 Gag protein. ELISA microtiter plates were coated with 0.2 μ g of HIV-1_{SP2} p24-Gag protein per well overnight and washed four times; subsequently, blocking was done with PBS-0.2% Tween (Sigma) for 2 hours. After removal of the blocking solution, 100 μ l of diluted mouse serum was added. Sera were tested at 1/25 dilutions and by serial 3-fold dilutions, thereafter. Microtiter plates were washed four times and incubated with a secondary, peroxidase-coupled anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA plates were washed and 100 μ l of 3, 3', 5, 5'-tetramethyl benzidine (TMB; Pierce) was added per well. The optical density of each well was measured after 15 minutes. The titers reported are the reciprocal of the dilution of serum that gave a half-maximum optical density (O.D.). The ELISA results are presented in Table 8.

Table 8

Group	Inoculum (μ g)	Expression cassette	Sera - Week 4 ³	Sera - Week 6	Sera - Week 8
1	20	S ¹ - gag	98	455	551
2	2	S - gag	59	1408	227
3	0.	S - gag	29	186	61
4	0.02	S - gag	< 20	< 20	< 20
5	20	S - gag	67	n.a. ⁴	n.a.
6	2	S - gag	63	n.a.	n.a.
7	0.	S - gag	57	n.a.	n.a.
8	0.02	S - gag	< 20	n.a.	n.a.
9	20	N ² - gag	43	n.a.	n.a.
10	2	N - gag	< 20	n.a.	n.a.
11	0.	N - gag	< 20	n.a.	n.a.
12	0.02	N - gag	< 20	n.a.	n.a.

1 = synthetic gag expression cassette (SEQ ID NO: 4)

2 = native gag expression cassette (SEQ ID NO: 1)

3 = geometric mean antibody titer

4 = not applicable

The results of the mouse immunizations with plasmid-DNAs show that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression cassettes. Also, the second boost immunization induced a secondary immune response after two weeks (groups 1-3).

C. Cellular Immune Response

The frequency of specific cytotoxic T-lymphocytes (CTL) was evaluated by a standard chromium release assay of peptide pulsed Balb/c mouse CD4 cells. Gag expressing vaccinia virus infected CD-8 cells were used as a positive control (vvGag). Briefly, spleen cells (Effector cells, E) were obtained from the BALB/c mice immunized as described above (Table 8) were cultured, restimulated, and assayed for CTL activity against Gag

peptide-pulsed target cells as described (Doe, B., and Walker, C.M., *AIDS* 10(7):793-794, 1996). The HIV-1_{SP2} Gag peptide used was p7g SEQ ID NO:10. Cytotoxic activity was measured in a standard ⁵¹Cr release assay. Target (T) cells were cultured with effector (E) cells at various E:T ratios for 4 hours and the average cpm from duplicate wells was used to calculate percent specific ⁵¹Cr release. The results are presented in Table 9.

Cytotoxic T-cell (CTL) activity was measured in splenocytes recovered from the mice immunized with HIV Gag DNA (compare Effector column, Table 9, to immunization schedule, Table 8). Effector cells from the Gag DNA-immunized animals exhibited specific lysis of Gag p7g peptide-pulsed SV-BALB (MHC matched) targets cells indicative of a CTL response. Target cells that were peptide-pulsed and derived from an MHC-unmatched mouse strain (MC57) were not lysed (Table 9; MC/p7g).

Table 9

Table 9. Cytotoxic T-lymphocyte (CTL) responses in mice immunized with HIV-1 gag DNA					
		Percent specific lysis of target cells'			
Immunization	E:T	SVBALB none	SVBALB p7g	RMA p7g	
20 µg DNA gagmod	100:1	2	49	<1	
	30:1	3	30	<1	
	10:1	<1	14	<1	
2 µg DNA gagmod	100:1	2	37	<1	
	30:1	2	21	<1	
	10:1	<1	13	<1	
0.2 µg DNA gagmod	100:1	2	32	<1	
	30:1	3	25	<1	
	10:1	1	14	<1	
0.02 µg DNA gagmod	100:1	1	17	<1	
	30:1	1	16	<1	
	10:1	1	8	<1	
20 µg DNA gag native	100:1	2	49	<1	
	30:1	2	24	<1	
	10:1	1	12	<1	
2 µg DNA gag native	100:1	<1	18	<1	
	30:1	1	14	<1	
	10:1	1	7	<1	
0.2 µg DNA gag native	100:1	3	30	<1	
	30:1	3	17	<1	
	10:1	2	7	<1	
0.02 µg DNA gag native	100:1	4	2	<1	
	30:1	1	2	<1	
	10:1	1	2	<1	

representative results of two animals per DNA-dose; positive CTL responses are indicated by boxed data

representative results of two animals per DNA-dose;
positive CTL responses are indicated by boxed data

The results of the CTL assays show increased potency of synthetic Gag expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.

Example 9In vivo Immunization with Env polypeptidesA. Immunogenicity Study of US4 o-gp140 in Ras-3c Adjuvant System

5 Studies have been conducted using rabbits immunized with US4 o-gp140 purified as described above. Studies are also underway in animals to determine immunogenicity of US4 gp120, SF162 o-gp140 and SF162 gp120.

10 Two rabbits (#1 and #2) were immunized intramuscularly at 0, 4, 12 and 24 weeks with 50 µg of US4 o-gp140 in the Ribi™ adjuvant system (RAS-3c), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL, Ribi Immunochem, Hamilton, MT).

15 In each experiment described herein, o-gp140 can be native, mutated and/or modified. Antibody responses directed against the US4 o-gp140 protein were measured by ELISA. Results are shown in Table 10.

Table 10

Rabbit/sample	Approximate o-gp140 ELISA titer
pre-immunization	0
#1: post1 (0 week immuniz)	400
#1: post2 (4 week immuniz)	15,000
#1: post3 (12 week immuniz)	50,000
#1: post4 (24 week immuniz)	100,000
#2: post1 (0 week immuniz)	600
#2: post2 (4 week immuniz)	12,000
#2: post3 (12 week immuniz)	25,000
#2: post4 (24 week immuniz)	55,000

The avidities of antibodies directed against the US4 o-gp140 protein were measured in a similar ELISA format employing successive washes with increasing concentrations of ammonium isothiocyanate. Results are shown in Table 11.

Table 11

Time of sample	Approx. Antibody avidity (NH ₄ SCN Conc. in M)
pre-immunization	0.02
post1 (0 week immuniz)	1.8
post2 (4 week immuniz)	3.5
post3 (12 week immuniz)	5.5
post4 (24 week immuniz)	5.1

These results show that US4 o-gp140 is highly immunogenic and able to induce substantial antibody responses after only one or two immunizations.

5 B. Immunogenicity of US4 o-gp140 in MF59-based Adjuvants

Groups of 4 rabbits were immunized intramuscularly at 0, 4, 12 and 24 weeks with various doses of US4 o-gp140 protein in three different MF59-based adjuvants (MF59 is described in International Publication No. WO 90/14837 and typically contains 5% Squalene, 0.5% Tween 80, and 0.5% Span 85). Antibody titers were measured post-third by ELISA using SF2 gp120 to coat the plates. QHC is a quill-based adjuvant (IscoTek, Uppsala, Sweden). Results are shown in Table 12.

15 Table 12

Antigen dose (μ g)	Adjuvant	Anti-gp120 _{SF2} Ab GMT*
12.5	MF59	7231
25	MF59	8896
50	MF59	12822
12.5	MF59/MPL	24146
25	MF59/MPL	27199
50	MF59/MPL	23059
50	MF59/MPL/QHC	31759

*GMT = geometric mean titer

Thus, adjuvanted o-gp140 generated antigen-specific antibodies. Further, the antibodies were shown to increased in avidity over time.

30 C. Neutralizing Antibodies

Neutralizing antibodies post-third immunization were measured against HIV-1 SF2 in a T-cell line adapted virus

(TCLA) assay and against PBMC-grown HIV-1 variants SF2, SF162 and 119 using the CCR5+ CEMx174 LTR-GFP reporter cell line, 5.25 (provided by N. Landau, Salk Institute, San Diego, CA) as target cells. Results are shown in Table 13.

5

Table 13

Neutralizing antibody responses in rabbits immunized
with o-gp140.modUS4 protein

	Group	Animal	SF2	SF2	SF162	119
			TCLA*	PBMC*	PBMC*	PBMC*
10	Experiment 1					
	o-gp140/ Ras-3c 50 mg	217	>640	100%	49	17
		218	>640	96	37	29
15	Experiment 2					
	o-gp140/ MF59 50 mg	792	45	71	39	26
		793	50	87	26	4
		794	59	87	13	0
20		795	128	92	15	0
	o-gp140/ MF59 + MPL 50 mg	804	173	91	47	18
		805	134	93	28	4
		806	N.D.**	95	49	13
25		807	441	100	31	15
	o-gp140/MF59 + MPL + QHC 50 mg	808	465	98	46	40
		809	496	100	44	39
		810	>640	101	27	4
30		811	92	92	24	37

*TCLA neutralizing antibody titers (50% inhibition).

**Not Determined

* % Inhibition at 1:10 dilution of sera with any detectable non-specific inhibition in pre-bleeds subtracted.

35

The above studies in rabbits indicate that the US4 o-gp140 protein is highly immunogenic. When administered with adjuvant, this protein was able to induce substantial antibody responses after only one or two immunizations. 5 Moreover, the adjuvanted o-gp140 protein was able to generate antigen-specific antibodies which increased in avidity after successive immunizations, and substantial neutralizing activity against T-cell line adapted HIV-1. Neutralizing activity was also observed against PBMC-grown 10 primary HIV strains, including the difficult to neutralize CCR5 co-receptor (R5)-utilizing isolates, SF162 and 119.

Example 10

In Vivo Immunogenicity of Synthetic Env Expression

15

Cassettes

A. General Immunization Methods

To evaluate the immunogenicity of the synthetic Env expression cassettes, studies using guinea pigs, rabbits, mice, rhesus macaques and baboons were performed. The studies were structured as follows: DNA immunization alone (single or multiple); DNA immunization followed by protein immunization (boost); DNA immunization followed by Sindbis particle immunization; immunization by Sindbis particles alone. 20

B. Humoral Immune Response

The humoral immune response was checked in serum specimens from immunized animals with an anti-HIV Env antibody ELISAs (enzyme-linked immunosorbent assays) at various times post-immunization. The antibody titers of the sera were determined by anti-Env antibody ELISA as described above. Briefly, sera from immunized animals were 30

screened for antibodies directed against the HIV gp120 or gp140 Env protein. Wells of ELISA microtiter plates were coated

overnight with the selected Env protein and washed four times; subsequently, blocking was done with PBS-0.2% Tween (Sigma) for 2 hours. After removal of the blocking solution, 100 μ l of diluted mouse serum was added. Sera were tested at 1/25 dilutions and by serial 3-fold dilutions, thereafter. Microtiter plates were washed four times and incubated with a secondary, peroxidase-coupled anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA plates were washed and 100 μ l of 3, 3', 5, 5'-tetramethyl benzidine (TMB; Pierce) was added per well. The optical density of each well was measured after 15 minutes. Titers are typically reported as the reciprocal of the dilution of serum that gave a half-maximum optical density (O.D.).

Example 11

DNA-immunization of Baboons Using Synthetic Gag

Expression Cassettes

A. Baboons

Four baboons were immunized 3 times (weeks 0, 4 and 8) bilaterally, intramuscular into the quadriceps using 1mg pCMVKM2.GagMod.SF2 plasmid-DNA (Example 1). The animals were bled two weeks after each immunization and a p24 antibody ELISA was performed with isolated plasma. The ELISA was performed essentially as described in Example 5 except the second antibody-conjugate was an anti-human IgG, g-chain specific, peroxidase conjugate (Sigma Chemical Co., St. Louis, MD 63178) used at a dilution of 1:500. Fifty μ g/ml yeast extract was added to the dilutions of plasma

samples and antibody conjugate to reduce non-specific background due to

preexisting yeast antibodies in the baboons. The antibody titer results are presented in Table 14.

5

Table 14

Immunization no.	Weeks	Antigen	wpi ^a / Baboon No.	Ab-titer ^b
10	0	gagmod DNA	0 w/219	< 10
			0 w/220	< 10
			0 w/221	< 10
			0 w/222	< 10
15	6		2 wp 1st/219	< 10
			2 wp 1st/220	< 10
			2 wp 1st/221	< 10
			2 wp 1st/222	15
20	14	gagmod DNA	2 wp 4th/219	< 10
			2 wp 4th/220	88
			2 wp 4th/221	< 10
			2 wp 4th/222	56
25	30	gagmod DNA	2 wp 5th/219	< 10
			2 wp 5th/220	391
			2 wp 5th/221	237
			2 wp 5th/222	222
30	46	gag VLP protein	2 wp 6th/219	753
			2 wp 6th/219	4330
			2 wp 6th/219	5000
			2 wp 6th/219	2881

^a wpi = weeks post immunization

^b geometric mean antibody titer

30

In Table 14, pre-bleed data are given as Immunization No. 0; data for bleeds taken 2 weeks post-first immunization are given as Immunization No. 1; data for bleeds taken 2 weeks post-second immunization are given as Immunization No. 2; and, data for bleeds taken 2 weeks post-third immunization are given as Immunization No. 3.

Further, lymphoproliferative responses to p24 antigen were also observed in baboons 221 and 222 two weeks post-fourth immunization (at week 14), and enhanced substantially post-boosting with VLP (at week 44 and 76).
5 Such proliferation results are indicative of induction of T-helper cell functions.

B. Rhesus Macaques

The improved potency of the codon-modified *gag* expression plasmid observed in mouse and baboon studies was confirmed in rhesus macaques. Four of four macaques had detectable Gag-specific CTL after two or three 1 mg doses of modified *gag* plasmid. In contrast, in a previous study, only one of four macaques given 1 mg
15 doses of plasmid-DNA encoding the wild-type HIV-1_{SF2} Gag showed strong CTL activity that was not apparent until after the seventh immunization. Further evidence of the potency of the modified *gag* plasmid was the observation that CTL from two of the four rhesus macaques reacted
20 with three nonoverlapping Gag peptide pools, suggesting that as many as three different Gag peptides are recognized and indicating that the CTL response is polyclonal. Additional quantification and specificity studies are in progress to further characterize the T
25 cell responses to Gag in the plasmid-immunized rhesus macaques. DNA immunization of macaques with the modified *gag* plasmid did not result in significant antibody responses, with only two of four animals seroconverting at low titers. In contrast, in the same study the
30 majority of macaques in groups immunized with p55Gag protein seroconverted and had strong Gag-specific antibody titers. These data suggest that a prime-boost

strategy (DNA-prime and protein-boost) could be very promising for the induction of a strong CTL and antibody response.

- 5 In sum, these results demonstrate that the synthetic Gag plasmid DNA is immunogenic in non-human primates. When similar experiments were carried out using wild-type Gag plasmid DNA no such induction of anti-p24 antibodies was observed after four immunizations.

10

Example 12

DNA- and Protein Immunizations of Animals Using Env Expression Cassettes and Polypeptides

A. Guinea Pigs

- Groups comprising six guinea pigs each were
15 immunized intramuscularly at 0, 4, and 12 weeks with plasmid DNAs encoding the gp120.modUS4, gp140.modUS4, gp140.modUS4.delV1, gp140.modUS4.delV2, gp140.modUS4.delV1/V2, or gp160.modUS4 coding sequences of the US4-derived Env. The animals were subsequently
20 boosted at 18 weeks with a single intramuscular dose of US4 o-gp140.mut.modUS4 protein in MF59 adjuvant. Anti-gp120 SF2 antibody titers (geometric mean titers) were measured at two weeks following the third DNA
immunization and at two weeks after the protein boost.
25 Results are shown in Table 15.

Table 15

Group	GMT post-DNA immuniz.	GMT post-protein boost
gp120.modUS4	2098	9489
gp140.modUS4	190	5340
gp140.modUS4.delV1	341	7808
gp140.modUS4.delV2	386	8165
gp140.modUS4.delV1/V 2	664	8270
gp160.modUS4	235	9928

These results demonstrate the usefulness of the synthetic constructs to generate immune responses, as well as, the advantage of providing a protein boost to enhance the immune response following DNA immunization.

E. Rabbits

Rabbits were immunized intramuscularly and intradermally using a Bioject needleless syringe with plasmid DNAs encoding the following synthetic SF162 Env polypeptides: gp120.modSF162, gp120.modSF162.delV2, gp140.modSF162, gp140.modSF162.delV2, gp140.mut.modSF162, gp140.mut.modSF162.delV2, gp160.modSF162, and gp160.modSF162.delV2. Approximately 1 mg of plasmid DNA (pCMVlink) carrying the synthetic Env expression cassette was used to immunize the rabbits. Rabbits were immunized with plasmid DNA at 0, 4, and 12 weeks. At two weeks after the third immunization all of the constructs were shown to have generated significant antibody titers in the test animals. Further, rabbits immunized with constructs containing deletions of the V2 region

generally generated similar antibody titers relative to rabbits immunized with the companion construct still containing the V2 region.

The nucleic acid immunizations are followed by protein boosting with o-gp140.modSF162.delV2 (0.1 mg of purified protein) at 24 weeks after the initial immunization. Results are shown in Table 16.

Table 16

Group	GMT 2wks post-2nd DNA immunization	GMT 2wks post-3rd DNA immunization	GMT 2wks post-protein boost
gp120.modSF162	4573	5899	26033
gp120.modSF162.delV2	3811	3122	29606
gp140.modSF162	1478	710	12882
gp140.modSF162.delV2	1572	819	11067
gp140.mut.modSF162	1417	710	8827
gp140.mut.modSF162.delV2	1378	1207	13301
gp160.modSF162	23	81	7050
gp160.modSF162.delV2	85	459	11568

All constructs are highly immunogenic and generate substantial antigen binding antibody responses after only 2 immunizations in rabbits.

C. Baboons

Groups of four baboons were immunized intramuscularly with 1 mg doses of DNA encoding different forms of synthetic US4 gp140 (see the following table) at 0, 4, 8, 12, 28, and 44 weeks. The animals were also boosted twice with US4 0-gp140 protein (gp140.mut.modUS4) at 44 and 76 weeks using MF59 as adjuvant. Results are shown in Table 17.

Table 17				
Animal	Treatment	2 Wks Post 5th DNA immuniza- tion	2 Wks post 6th DNA (plus o- gp140 prot. immuniz.)	2 Wks post 7th DNA (o-gp140 protein only)
5 CY 215	gp140.modUS4	8.3	446	1813
CY 216		8.3	433	1236
CY 217		68	1660	2989
CY 218		101	2556	1610
Geomean:		26.2	951.4	1812.1
10 CY 219	gp140.modUS4 + p55gag.SF2	8.3	8.3	421
CY 220		8.3	8.3	3117
CY 221		8.3	954	871
CY 222		8.3	71	916
Geomean:		8.3	46.5	1011.5
15 CY 223	gp140.mut. modUS4	41.4	10497	46432
CY 224		8.3	979	470
CY 225		135	2935	3870
CY 226		47	1209	4009
Geomean:		68.3	2457.4	4289.6
20 CY 227	gp140TM. modUS4	8.3	56	5001
CY 228		8.3	806	1170
CY 229		8.3	48	3402
CY 230		8.3	38	6520
GMT*:		8.3	95.3	3375.3

*GMT = geometric mean titer

The results in Table 17 demonstrate the usefulness of the synthetic constructs to generate immune responses in primates such as baboons. In addition, all animals

showed evidence of antigen-specific (*Env* antigen) lymphoproliferative responses.

D. Rhesus Macaques

5 Two rhesus macaques (designated H445 and J408) were immunized with 1 mg of DNA encoding SF162 gp140 with a deleted V2 region (SF162.gp140.delV2) by intramuscular (IM) and intradermal (ID) routes at 0, 4, 8, and 28 weeks. Approximately 100 µg of the protein encoded by
10 the SF162. gp140mut.delV2 construct was also administered in MF59 by IM delivery at 28 weeks.

ELISA titers are shown in Figure 61. Neutralizing antibody activity is shown Tables 18 and 19. Neutralizing antibody activity was determined against a
15 variety of primary HIV-1 isolates in a primary lymphocyte or "PBMC-based" assay (see the following tables). Further, the phenotypic co-receptor usage for each of the primary isolates is indicated. As can be seen in the
tables neutralizing antibodies were detected against
20 every isolate tested, including the HIV-1 primary isolates (i.e., SF128A, 92US660, 92HT593, 92US657, 92US714, 91US056, and 91US054).

Table 18					
	Treatment		Bleed 0	Bleed 1	Bleed 2
Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
5	EO 456	(None)	8.3	45	309
	EO 457		8.3	254	460
	EO 458		8.3	8.3	93
	EO 459		8.3	43	45
	EO 460		8.3	8.3	274
10	EO 461	25 μ g 120mod DNA	8.3	47	1502
	EO 462		8.3	80	5776
	EO 463		8.3	89	3440
	EO 464		8.3	8.3	3347
	EO 465		8.3	69	1127
15	EO 466	(None)	8.3	63	102
	EO 467		8.3	112	662
	EO 468		8.3	94	459
	EO 469		8.3	58	48
	EO 470		8.3	95	355
20	EO 471	50 μ g 120mod DNA	8.3	110	9074
	EO 472		8.3	8.3	4897
	EO 473		8.3	49	4089
	EO 474		8.3	59	5280
	EO 475		8.3	8.3	929
25	EO 476	Sindbis/Env	8.3		653
	EO 477		8.3	87	22675
	EO 478		8.3	76	3869
	EO 479		8.3		1004
	EO 480		8.3	71	7080

Table 19					
	Treatment		Bleed 0	Bleed 1	Bleed 2
Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
EO 481	Sindbis/Env	(None)	8.3	8.3	8.3
EO 482			8.3	8.3	8.3
EO 483			8.3	78	103
EO 484			8.3	8.3	32
EO 485			8.3	76	207
EO 486	Sindbis/Env	Sindbis/Env	8.3	8.3	458
EO 487			8.3	8.3	345
EO 488			8.3	8.3	331
EO 489			8.3	103	111
EO 490			8.3	8.3	5636

Lymphoproliferative activity (LPA) was also determined by antigenic stimulation followed by uptake of ³H-thymidine in these animals and is shown in Table 20. Experiment 1 was performed at 14 weeks post third DNA immunization and Experiment 2 was performed at 2 weeks post fourth DNA immunization using DNA and protein. For gp120ThaiE, gp120SF2 and US4 o-gp140, appropriate background values were used to calculate Stimulation Indices (S.I.; Antigenic stimulation CPM/Background CPM).

Table 20

S.I.: Calculated as Ag CPM/Background CPM				
Animal/ exp#	gp120Thai E	gp120 SF2	env2-3SF2	o- gp140US4
J408/#1	2	1	1	5
H445/#1	1	1	1	6
J408/#2	1	1	2	3
H445/#2	0	0	3	2

As can be seen by the results presented in Table 20 lymphoproliferative responses to o-gp140.US4 antigen were also in all four animals at both experimental time points. Such proliferation results are indicative of induction of T-helper cell functions.

The results presented above demonstrate that the synthetic gp140.modSF162.delV2 DNA and protein are immunogenic in non-human primates.

Example 13

In vitro expression of recombinant Sindbis RNA and
DNA containing the synthetic Gag or Env expression
cassettes

5 A. Synthetic Gag expression cassettes

To evaluate the expression efficiency of the synthetic Gag expression cassette in Alphavirus vectors, the synthetic Gag expression cassette was subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors. Specifically, a cDNA vector construct for *in vitro* transcription of Sindbis virus RNA vector replicons (pRSIN-luc; Dubensky, et al., *J Virol.* 70:508-519, 1996) was modified to contain a *PmeI* site for plasmid linearization and a polylinker for insertion of heterologous genes. A polylinker was generated using two oligonucleotides that contain the sites *XhoI*, *PmlI*, *ApaI*, *NarI*, *XbaI*, and *NotI* (XPANXNF, SEQ ID NO:17, and XPANXNR, SEQ ID NO:18).

The plasmid pRSIN-luc (Dubensky et al., *supra*) was digested with *XhoI* and *NotI* to remove the luciferase gene insert, blunt-ended using Klenow and dNTPs, and purified from an agarose gel using GeneCleanII (Biol01, Vista, CA). The oligonucleotides were annealed to each other and ligated into the plasmid. The resulting construct was digested with *NotI* and *SacI* to remove the minimal Sindbis 3'-end sequence and A₄₀ tract, and ligated with an approximately 0.4 kbp fragment from PKSSIN1-BV (WO 97/38087). This 0.4 kbp fragment was obtained by digestion of pKSSIN1-BV with *NotI* and *SacI*, and purification after size fractionation from an agarose gel. The fragment contained the complete Sindbis virus 3'-end, an A₄₀ tract and a *PmeI* site for linearization. This new vector construct was designated SINEVE.

The synthetic HIV Gag coding sequence was obtained from the parental plasmid by digestion with *EcoRI*, blunt-ending with Klenow and dNTPs, purification with GeneCleanII, digestion with *SalI*, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic Gag coding fragment was ligated into the SINEVE vector that had been digested with *XhoI* and *PmlI*. The resulting vector was purified using GeneCleanII and designated SINEVGag. Vector RNA replicons may be transcribed in vitro (Dubensky et al., *supra*) from SINEVGag and used directly for transfection of cells. Alternatively, the replicons may be packaged into recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line as described, for example, in U.S. Patent Numbers 5,843,723 and 5,789,245, and then administered in vivo as described..

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) was digested with *SalI* and *XbaI*, to remove the beta-galactosidase gene insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Gag gene was inserted into the the pDCMVSIN-beta-gal by digestion of SINEVGag with *SalI* and *XhoI*, purification using GeneCleanII of the Gag-containing fragment after agarose gel size fractionation, and ligation. The resulting construct was designated pDSIN-Gag, and may be used directly for *in vivo* administration or formulated using any of the methods described herein.

BHK and 293 cells were transfected with recombinant Sindbis vector RNA and DNA, respectively. The supernatants and cell lysates were tested with the Coulter p24 capture ELISA (Example 2).

BHK cells were transfected by electroporation with recombinant Sindbis RNA. The expression of p24 (in ng/ml) is presented in Table 21. In the table, SINGag#1 and 2 represent duplicate measurements, and SIN β gal represents a negative control. Supernatants and lysates were collected 24h post transfection.

Table 21

Construct	Supernatant	Lysate
SIN β gal RNA	0	0
SINGag#1 RNA	7 ng	Max (approx. 1 μ g)
SINGag#2 RNA	1 ng	700 ng

293 cells were transfected using LT-1 (Example 2) with recombinant Sindbis DNA. Synthetic pCMVKM2GagMod.SF2 was used as a positive control. Supernatants and lysates were collected 48h post transfection. The expression of p24 (in ng/ml) is presented in Table 22.

Table 22

Construct	Supernatant	Lysate
SINGag DNA	3	30
pCMVKM2.GagMod.SF2 DNA	32	42

The results presented in Tables 21 and 22 demonstrate that Gag proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector systems using the synthetic Gag expression cassette (p55Gag.mod).

B. Synthetic Env expression cassettes

To evaluate the expression efficiency of the synthetic Env expression cassette in Alphavirus vectors,

synthetic Env expression cassettes were subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors as described above for Gag.

The synthetic HIV Env coding sequence was obtained from the parental plasmid by digestion with *SalI* and *XbaI*, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic Env coding fragment was ligated into the SINBVE vector that had been digested with *XhoI* and *XbaI*. The resulting vector was purified using GeneCleanII and designated SINBVEEnv. Vector RNA replicons may be transcribed *in vitro* (Dubensky et al., *supra*) from SINBVEEnv and used directly for transfection of cells. Alternatively, the replicons may be packaged into recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line and administered as described above for Gag.

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) was digested with *SalI* and *XbaI*, to remove the beta-galactosidase gene insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Env gene was inserted into the the pDCMVSIN-beta-gal by digestion of SINBVEEnv with *XbaI* and *XhoI*, purification using GeneCleanII of the Env-containing fragment after agarose gel size fractionation, and ligation. The resulting construct was designated pDSIN-Env, and may be used directly for *in vivo* administration or formulated using any of the methods described herein.

BHK and 293 cells were transfected with recombinant Sindbis vector RNA and DNA, respectively. The supernatants and cell lysates were tested by capture ELISA.

BHK cells were transfected by electroporation with recombinant Sindbis RNA. The expression of Env (in ng/ml) is presented in Table 23. In the table, the Sindbis RNA containing synthetic Env expression cassettes are indicated and β gal represents a negative control. Supernatants and lysates were collected 24h post transfection.

Table 23

Construct	Supernatant (Neat)ng/ml	Lysate (1:10 dilution)ng/ml
β gal RNA	0	0
gp140.modUS4	726	7147
gp140.modSF162	3529	7772
gp140.modUS4.delV1/V2	1738	6526
gp140.modUS4.delV2	960	3023
gp140.modSF162.delV2	2772	3359

293 cells were transfected using LT-1 mediated transfection (PanVera) with recombinant Sindbis DNA containing synthetic expression cassettes of the present invention and β gal sequences as a negative control. Supernatants and lysates were collected 48h post transfection. The expression of Env (in ng/ml) is presented in Table 24.

Table 24

Construct	Supernatant (Neat)ng/ml	Lysate (1:10 dilution)ng/ml
8gal	0	0
gp140.modSF162.delV2	1977	801
gp140.modSF162	949	746

The results presented in Tables 23 and 24 demonstrated that Env proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector systems using the synthetic Env expression cassettes of the present invention.

Example 14

A. In vivo Immunization with Gag-containing DNA and/or Sindbis particles

CB6F1 mice were immunized intramuscularly at 0 and 4 weeks with plasmid DNA and/or Sindbis vector RNA-containing particles each containing GagMod.SF2 sequences as indicated in Table 25. Animals were challenged with recombinant vaccinia expressing SF2 Gag at 3 weeks post second immunization (at week 7). Spleens were removed from the immunized and challenged animals 5 days later for a standard ⁵¹C release assay for CTL activity. Values shown in Table 25 indicate the results from the spleens of three mice from each group. The boxed values in Table 25 indicate that all groups of mice receiving immunizations with pCMVKm2.GagMod.SF2 DNA and/or SindbisGagMod.SF2 virus particles either alone or in combinations showed antigen-specific CTL activity.

Table 25

Cytotoxic T-lymphocyte (CTL) responses in mice immunized with HIV-1 gagmod DNA and Sindbis gagmod virus particles					
Immunization	E:T	Percent specific lysis of target cells*			RMA
		SVBALB	SVBALB	SVBALB	
		none	p7g	p7g	
pCMVKm2.GagMod.SF2 DNA ^a	100:1	5	20	1	
at 0, 4 wks	25:1	5	20	<1	
	6:1	4	8	<1	
SindbisGagMod.SF2	100:1	10	49	<1	
virus particles ^b	25:1	7	20	<1	
at 0, 4 weeks	6:1	5	12	<1	
pCMVKm2.GagMod.SF2 DNA at 0	100:1	9	58	<1	
wks SindbisGagMod.SF2 virus	25:1	7	42	2	
particles at 4 wks	6:1	4	13	<1	
SindbisGagMod.SF2	100:1	5	38	<1	
virus particles at 4 wks	25:1	4	18	<1	
pCMVKm2.GagMod.SF2 DNA at 0 wks	6:1	3	13	1	

^a 20 µg^b 10⁷ particles

* Challenge with recombinant vaccinia virus expressing HIV-1SF2 Gag at 3 weeks post second immunization (week 7). Spleens taken 5 days later. Ex vivo CTL assay performed by standard ⁵¹Cr release assay. Values seen represent results from 3 pooled mouse spleens per group

B. In vivo Immunization with Env-containing DNA and/or Sindbis particles

Balb/C mice were immunized intramuscularly at 0 and 4 weeks (as shown in the following table) with plasmid DNA and/or Sindbis-virus RNA-containing particles each containing gp120.modUS4 sequences. Treatment regimes and antibody titers are shown in Table 26. Antibody titers were determined by ELISA using gp120 SF2 protein to coat the plates.

Table 26					
	Treatment		Bleed 0	Bleed 1 (8 wks)	Bleed 2 (10 wks)
Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
5	EO 456 EO 457 EO 458 EO 459 EO 460	25µg 120mod DNA (None)	8.3 8.3 8.3 8.3 8.3	45 254 8.3 43 8.3	309 460 93 45 274
10	EO 461 EO 462 EO 463 EO 464 EO 465	25µg 120mod DNA 25µg 120mod DNA	8.3 8.3 8.3 8.3 8.3	47 80 89 8.3 69	1502 5776 3440 3347 1127
15	EO 466 EO 467 EO 468 EO 469 EO 470	50µg 120mod DNA (None)	8.3 8.3 8.3 8.3 8.3	63 112 94 58 95	102 662 459 48 355
20	EO 471 EO 472 EO 473 EO 474 EO 475	50µg 120mod DNA 50µg 120mod DNA	8.3 8.3 8.3 8.3 8.3	110 8.3 49 59 8.3	9074 4897 4089 5280 929
25	EO 476 EO 477 EO 478 EO 479 EO 480	25µg 120mod DNA Sindbis/Env	8.3 8.3 8.3 8.3 8.3	87 76 71	653 22675 3869 1004 7080
30	EO 481 EO 482 EO 483 EO 484 EO 485	Sindbis/Env (None)	8.3 8.3 8.3 8.3 8.3	8.3 8.3 78 8.3 76	8.3 8.3 103 32 207
35	EO 486 EO 487 EO 488 EO 489 EO 490	Sindbis/Env Sindbis/Env	8.3 8.3 8.3 8.3 8.3	8.3 8.3 8.3 103 8.3	458 345 331 111 5636

As can be seen from the data presented above, all of the mice generally demonstrated substantial immunological responses by bleed number 2. For Env, the best results were obtained using either (i) 50 µg of gp120.modUS4 DNA for the first immunization followed by a second

immunization using 50 µg of gp120.modUS4 DNA, or (ii) 25 µg of gp120.modUS4 DNA for the first immunization followed by a second immunization using 10⁷ pfus of Sindbis.

- 5 The results presented above demonstrate that the Env and Gag proteins of the present invention are effective to induce an immune response using Sindbis vector systems which include the synthetic Env (e.g., gp120.modUS4) or Gag expression cassettes.

10

Example 15

Co-Transfection of Env and Gag as Monocistronic and Bicistronic Constructs

- DNA constructs encoding (i) wild-type US4 and SF162 Env polypeptides, (ii) synthetic US4 and SF162 Env polypeptides (gp160.modUS4, gp160.modUS4.delV1/V2, gp160.modSF162, and gp120.modSF162.delV2), and (iii) SF2gag polypeptide (i.e., the Gag coding sequences obtained from the SF2 variant or optimized sequences corresponding to the gagSF2 -- gag.modSF2) were prepared. These monocistronic constructs were co-transfected into 293T cells in a transient transfection protocol using the following combinations: gp160.modUS4; gp160.modUS4 and gag.modSF2; gp160.modUS4.delV1/V2; gp160.modUS4.delV1/V2 and gag.modSF2; gp160.modSF162 and gag.modSF2; gp120.modSF162.delV2 and gag.modSF2; and gag.modSF2 alone.
- 15
- 20
- 25

- Further several bicistronic constructs were made where the coding sequences for Env and Gag were under the control of a single CMV promoter and, between the two coding sequences, an IRES (internal ribosome entry site (EMCV IRES); Kozak, M., Critical Reviews in Biochemistry and Molecular Biology 27(45):385-402, 1992; Witherell, G.W., et al., Virology 214:660-663, 1995) sequence was
- 30

introduced after the Env coding sequence and before the Gag coding sequence. Those constructs were as follows:
gp160.modUS4.gag.modSF2, SEQ ID NO:73 (Figure 61);
gp160.modUSF162.gag.modSF2, SEQ ID NO:74 (Figure 62);
5 gp160.modUS4.delV1/V2.gag.modSF2, SEQ ID NO:75 (Figure 63); and gp160.modSF162.delV2.gag.modSF2, SEQ ID NO:76 (Figure 64).

Supernatants from cell culture were filtered through 0.45 μ m filters then ultracentrifuged for 2 hours at
10 24,000 rpm (140,000Xg) in an SW28 rotor through a 20% sucrose cushion. The pelleted materials were suspended and layered on a 20-60% sucrose gradient and spun for 2 hours at 40,000 rpm (285,000Xg) in an SW41Ti rotor. Gradients were fractionated into 1.0 ml samples. A total
15 of 9-10 fractions were typically collected from each DNA transfection group.

The fractions were tested for the presence of the Env and Gag proteins (across all fractions). These results demonstrated that the appropriate proteins were
20 expressed in the transfected cells (i.e., if an Env coding sequence was present the corresponding Env protein was detected; if a Gag coding sequence was present the corresponding Gag protein was detected).

Virus like particles (VLPs) were known to be present
25 through a selected range of sucrose densities. Chimeric virus like particles (VLPs) were formed using all the tested combinations of constructs containing both Env and Gag. Significantly more protein was found in the supernatant collected from the cells transfected with
30 "gp160.modUS4.delV1/V2 and gag.modSF2" than in all the other supernatants.

Western blot analysis was also performed on sucrose gradient fractions from each transfection. The results show that bicistronic plasmids gave lower amounts of VLPs

than the amounts obtained using co-transfection with monocistronic plasmids.

In order to verify the production of chimeric VLPs by these cell lines the following electron microscopic analysis was carried out.

293T cells were plated at a density of 60-70% confluence in 100 mm dishes on the day before transfection. The cells were transfected with 10 µg of DNA in transfection reagent LT1 (Panvera Corporation, 545 Science Dr., Madison, WI). The cells were incubated overnight in reduced serum medium (opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium was replaced with 10% fetal calf serum, 2% glutamine in IMDM in the morning of the next day and the cells were incubated for 65 hours. Supernatants and lysates were collected for analysis as described above (see Example 2).

The fixed, transfected 293T cells and purified ENV-GAG VLPs were analyzed by electron microscopy. The cells were fixed as follows. Cell monolayers were washed twice with PBS and fixed with 2% glutaraldehyde. For purified VLPs, gradient peak fractions were collected and concentrated by ultracentrifugation (24,000 rpm) for 2 hours. Electron microscopic analysis was performed by Prof. T.S. Benedict Yen (Veterans Affairs, Medical Center, San Francisco, CA).

Electron microscopy was carried out using a transmission electron microscope (Zeiss 10c). The cells were pre-stained with osmium and stained with uranium acetate and lead citrate. Immunostaining was performed to visualize envelope on the VLP. The magnification was 100,000X.

Figures 65A-65F show micrographs of 293T cells transfected with the following constructs: Figure 65A, gag.modSF2; Figure 65B, gpl60.modUS4; Figure 65C,

gpl60.modUS4.delV1/V2.gag.modSF2 (bicistronic Env and Gag); Figures 65D and 65E, gpl60.modUS4.delV1/V2 and gag.modSF2; and Figure 65F, gpl20.modSF162.delV2 and gag.modSF2. In the figures, free and budding immature virus-like-particles (VLPs) of the expected size (approximately 100 nm) decorated with the Env protein were seen. In sum, gp160 polypeptides incorporate into Gag VLPs when constructs were co-transfected into cells. The efficiency of incorporation is 2-3 fold higher when constructs encoding V-deleted Env polypeptides from high synthetic expression cassettes are used.

Although preferred embodiments of the subject invention have been described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the invention as defined by the appended claims.

What Is Claimed Is:

1. An expression cassette, comprising
5 a polynucleotide sequence encoding a polypeptide including an HIV Gag polypeptide, wherein the polynucleotide sequence encoding said Gag polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20.
10
2. The expression cassette of claim 1, comprising, a polynucleotide sequence encoding a polypeptide including an HIV Gag polypeptide, wherein the polynucleotide sequence encoding said Gag polypeptide
15 comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:9.
3. The expression cassette of claim 1, wherein said polynucleotide sequence encoding a polypeptide including
20 an HIV Gag polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:4.
4. The expression cassette of claim 1, wherein said
25 polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *protease* polypeptide.
5. The expression cassette of claim 4, wherein the nucleotide sequence encoding said polypeptide comprises a
30 sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:5, SEQ ID NO:78, and SEQ ID NO:79.
6. The expression cassette of claim 1, wherein said

polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *reverse transcriptase* polypeptide.

5 7. The expression cassette of claim 6, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, and SEQ
10 ID NO:84.

 8. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *tat* polypeptide.

15 9. The expression cassette of claim 8, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID
20 NO:86, SEQ ID NO:87, SEQ ID NO:88 and SEQ ID NO:89.

 10. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a
25 polynucleotide sequence encoding an HIV *polymerase* polypeptide, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID
NO:6.

30 11. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *polymerase* polypeptide, wherein (i) the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90%

sequence identity to the sequence presented as SEQ ID NO:4, and (ii) wherein the sequence is modified by deletions of coding regions corresponding to reverse transcriptase and integrase.

5

12. The expression cassette of claim 11, wherein said polynucleotide sequence preserves T-helper cell and CTL epitopes.

10

13. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HCV core polypeptide, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:7.

15

14. An expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV Env polypeptide, wherein the polynucleotide sequence encoding said Env polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59).

20

15. The expression cassette of claim 14, wherein said Env polypeptide includes sequences flanking a V1 region but has a deletion in the V1 region itself.

25

16. The expression cassette of claim 15, wherein the polynucleotide sequence encoding the polypeptide comprises the sequence presented as SEQ ID NO:65 (Figure 52 gpl60.modUS4.delV1).

30

17. The expression cassette of claim 14, wherein

said *Env* polypeptide includes sequences flanking a V2 region but has a deletion in the V2 region itself.

18. The expression cassette of claim 17, wherein
5 the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:60 (Figure 47); and SEQ ID NO:66 (Figure 53).

19. The expression cassette of claim 17, wherein
10 the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:34 (Figure 20); SEQ ID NO:37 (Figure 24); SEQ ID NO:40 (Figure 27); SEQ ID NO:43 (Figure 30); SEQ ID NO:46 (Figure 33); SEQ ID NO:49 (Figure 36); and SEQ ID NO:76
15 (Figure 64).

20. The expression cassette of claim 14, wherein
said *Env* polypeptide includes sequences flanking a V1/V2 region but has a deletion in the V1/V2 region itself.
20

21. The expression cassette of claim 20, wherein
the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:59 (Figure 46); SEQ ID NO:61 (Figure 48); SEQ ID NO:67
25 (Figure 54); and SEQ ID NO:75 (Figure 63).

22. The expression cassette of claim 20, wherein
the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:35
30 (Figure 21); SEQ ID NO:38 (Figure 25); SEQ ID NO:41 (Figure 28); SEQ ID NO:44 (Figure 31); SEQ ID NO:47 (Figure 34) and SEQ ID NO:50 (Figure 37).

23. The expression cassette of claim 14, wherein said *Env* polypeptide has a mutated cleavage site that prevents the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide.

5

24. The expression cassette of claim 23, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:57 (Figure 44); SEQ ID NO:61 (Figure 48); and SEQ ID NO:63 (Figure 50).

10

25. The expression cassette of claim 23, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34).

15

20

26. The expression cassette of claim 14, wherein said *Env* polypeptide includes a gp160 *Env* polypeptide or a polypeptide derived from a gp160 *Env* polypeptide.

25

27. The expression cassette of claim 26, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:64 (Figure 51); SEQ ID NO:65 (Figure 52); SEQ ID NO:66 (Figure 53); SEQ ID NO:67 (Figure 54); SEQ ID NO:68 (Figure 55); SEQ ID NO:75 (Figure 63); and SEQ ID NO:73 (Figure 61).

30

28. The expression cassette of claim 26, wherein

the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:48 (Figure 35); SEQ ID NO:49 (Figure 36); SEQ ID NO:50 (Figure 37); SEQ ID NO:76 (Figure 64); and SEQ ID NO:74 (Figure 62).

29. The expression cassette of claim 14, wherein said *Env* polypeptide includes a gp140 *Env* polypeptide or a polypeptide derived from a gp140 *Env* polypeptide.

30. The expression cassette of claim 29, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:56 (Figure 43); SEQ ID NO:57 (Figure 44); SEQ ID NO:58 (Figure 45); SEQ ID NO:59 (Figure 46); SEQ ID NO:60 (Figure 47); SEQ ID NO:61 (Figure 48); SEQ ID NO:62 (Figure 49); and SEQ ID NO:63 (Figure 50).

31. The expression cassette of claim 29, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:36 (Figure 23); SEQ ID NO:37 (Figure 24); SEQ ID NO:38 (Figure 25); SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34).

32. The expression cassette of claim 14, wherein said *Env* polypeptide includes a gp120 *Env* polypeptide or a polypeptide derived from a gp120 *Env* polypeptide.

33. The expression cassette of claim 32, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:54 (Figure 41); and SEQ ID NO:55 (Figure 42).

5

34. The expression cassette of claim 32, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:33 (Figure 19); SEQ ID NO:34 (Figure 20); and SEQ ID NO:35 (Figure 21).

10

35. The expression cassette of claim 14, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:55 (Figure 42); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); and SEQ ID NO:68 (Figure 55).

15

36. A recombinant expression system for use in a selected host cell, comprising, an expression cassette of any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the selected host cell.

20

37. The recombinant expression system of claim 36, wherein said control elements are selected from the group consisting of a transcription promoter, a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

25

38. The recombinant expression system of claim 36, wherein said transcription promoter is selected from the

group consisting of CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein.

39. A cell comprising an expression cassette of any
5 of claims 1-35, and wherein said polynucleotide sequence
is operably linked to control elements compatible with
expression in the selected cell.

40. The cell of claim 39, wherein the cell is a
10 mammalian cell.

41. The cell of claim 40, wherein the cell is
selected from the group consisting of BHK, VERO, HT1080,
293, RD, COS-7, and CHO cells.
15

42. The cell of claim 41, wherein said cell is a
CHO cell.

43. The cell of claim 39, wherein the cell is an
20 insect cell.

44. The cell of claim 43, wherein the cell is
either *Trichoplusia ni* (Tn5) or Sf9 insect cells.

45. The cell of claim 39, wherein the cell is a
25 bacterial cell.

46. The cell of claim 39, wherein the cell is a
yeast cell.
30

47. The cell of claim 39, wherein the cell is a
plant cell.

48. The cell of claim 39, wherein the cell is an antigen presenting cell.

49. The cell of claim 48, wherein the lymphoid cell
5 is selected from the group consisting of macrophage,
monocytes, dendritic cells, B-cells, T-cells, stem cells,
and progenitor cells thereof.

50. The cell of claim 39, wherein the cell is a
10 primary cell.

51. The cell of claim 39, wherein the cell is an immortalized cell.

52. The cell of claim 39, wherein the cell is a
15 tumor-derived cell.

53. A method for producing a polypeptide including
HIV *Gag* polypeptide sequences, said method comprising,
20 incubating the cells of claim 39, under conditions
for producing said polypeptide.

54. A method for producing virus-like particles
(VLPs), comprising,
25 incubating the cells of claim 39, under conditions
for producing said VLPs.

55. A method for producing a composition of virus-
like particles (VLPs), comprising,
30 (a) incubating the cells of claim 39, under
conditions for producing said VLPs; and
(b) substantially purifying said VLPs to produce a
composition of VLPs.

56. A cell line useful for packaging lentivirus vectors, comprising

suitable host cells that have been transfected with an expression vector containing an expression cassette of
5 any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the host cell.

57. The cell line of claim 56, wherein suitable
10 host cells have been transfected with an expression vector containing the expression cassette of any of claims 1-13.

58. The cell line of claim 56, wherein suitable
15 host cells have been transfected with an expression vector containing the expression cassette of claim 1-3.

59. The cell line of claim 56, wherein suitable
20 host cells have been transfected with an expression vector containing the expression cassette of claim 14-35.

60. A gene delivery vector for use in a Mammalian subject, comprising

a suitable gene delivery vector for use in said
25 subject, wherein the vector comprises an expression cassette of any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the subject.

30 61. A method of DNA immunization of a subject, comprising,

introducing a gene delivery vector of claim 60 into said subject under conditions that are compatible with expression of said expression cassette in said subject.

62. The method of claim 61, wherein said gene delivery vector is a nonviral vector.

5 63. The method of claim 61, wherein said vector is delivered using a particulate carrier.

64. The method of claim 63, wherein said vector is coated on a gold or tungsten particle and said coated
10 particle is delivered to said subject using a gene gun.

65. The method of claim 63, wherein said vector is encapsulated in a liposome preparation.

15 66. The method of claim 61, wherein said vector is a viral vector.

67. The method of claim 66, wherein said viral vector is a retroviral vector.
20

68. The method of claim 67, wherein said viral vector is a lentiviral vector.

69. The method of claim 61, wherein said subject is
25 a mammal.

70. The method of claim 69, wherein said mammal is a human.

30 71. A method of generating an immune response in a subject, comprising

transfecting cells of said subject a gene delivery vector of claim 60, under conditions that permit the expression of said polynucleotide and production of said

polypeptide, thereby eliciting an immunological response to said polypeptide.

72. The method of claim 71, wherein said vector is
5 a nonviral vector.

73. The method of claim 72, wherein said vector is delivered using a particulate carrier.

10 74. The method of claim 73, wherein said vector is coated on a gold or tungsten particle and said coated particle is delivered to said vertebrate cell using a gene gun.

15 75. The method of claim 73, wherein said vector is encapsulated in a liposome preparation.

76. The method of claim 71, wherein said vector is a viral vector.
20

77. The method of claim 76, wherein said viral vector is a retroviral vector.

25 78. The method of claim 77, wherein said viral vector is a lentiviral vector.

79. The method of claim 71, wherein said subject is a mammal.

30 80. The method of claim 79, wherein said mammal is a human.

81. The method of claim 71, wherein said transfecting is done ex vivo and said transfected cells

are reintroduced into said subject.

82. The method of claim 71, wherein said transfecting is done *in vivo* in said subject.

5

83. The method of claim 71, where said immune response is a humoral immune response.

84. The method of claim 71, where said immune response is a cellular immune response.

10

85. A gene delivery vector comprising an alphavirus vector construct, wherein said alphavirus construct comprises an expression cassette according to any one of claims 1 through 35.

15

86. The gene delivery vector of claim 85, wherein the alphavirus vector construct is a cDNA vector construct.

20

87. The gene delivery vector of claim 85, wherein the alphavirus comprises a recombinant alphavirus particle preparation.

88. The gene delivery vector of claim 85, wherein the vector comprises a eukaryotic layered vector initiation system.

25

89. A method of stimulating an immune response in a subject comprising administering the gene delivery vector of any one of claims 85 through 88 in an amount effective to stimulate an immune response in said subject.

30

90. The method of claim 89, wherein the gene

delivery vector is administered intramuscularly, intramucosally, intranasally, subcutaneously, intradermally, transdermall, intravaginally, intrarectally, orally or intravenously.

1 / 131

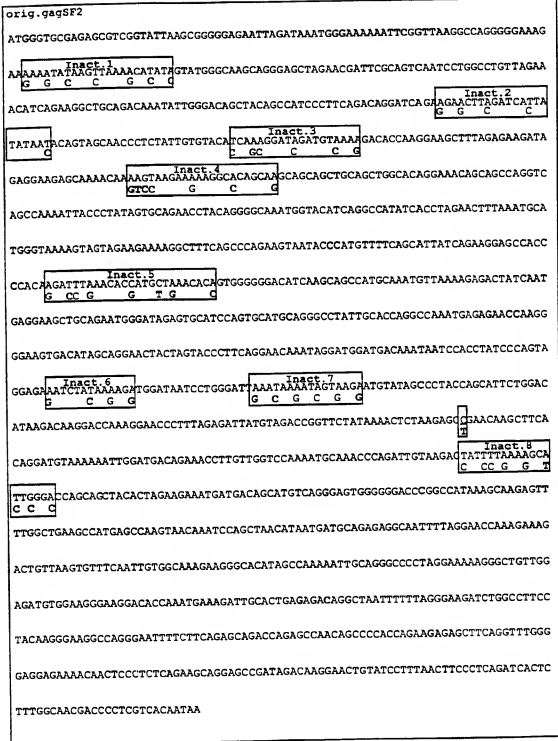


FIG. 1

native HIV-1SF2 gag-protease

2 / 131

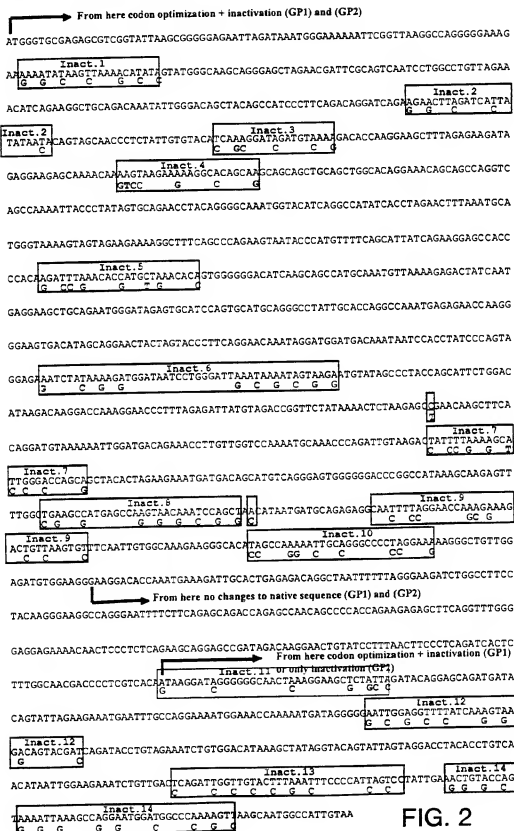


FIG. 2

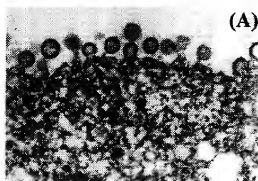


FIG. 3A

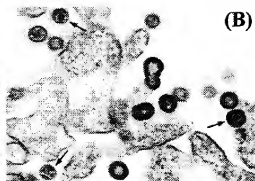


FIG. 3B

4 / 131

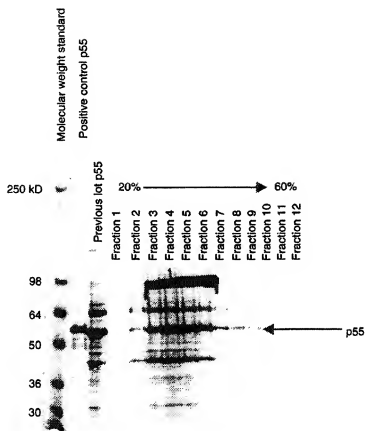


FIG. 4

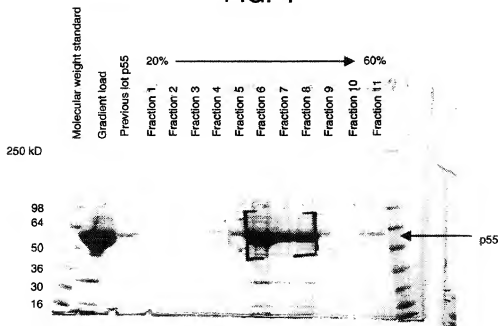


FIG. 5

5 / 131

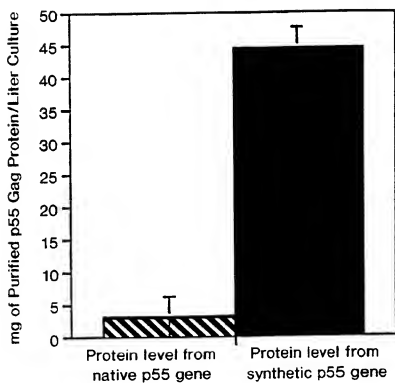


FIG. 6

6 / 131

FIG. 7A

GagPol.ModSF	1	ATGGGCGCC	10	20	30	40	50
GagProt.ModS	1	ATGGGCGCC	GCCCAGCGT	GCTGAGCGG	GGCGAGCTG	ACAAGTGGG	50
Gag.ModSF2	1	ATGGGCGCC	GCCCAGCGT	GCTGAGCGG	GGCGAGCTG	ACAAGTGGG	50
GagPol.ModSF	51	GAAGATCCG	CTGGGCCCC	GGCGGAGAA	GAAGTACAG	CTGAGACAA	100
GagProt.ModS	51	GAAGATCCG	CTGGGCCCC	GGCGGAGAA	GAAGTACAG	CTGAGACAA	100
Gag.ModSF2	51	GAAGATCCG	CTGGGCCCC	GGCGGAGAA	GAAGTACAG	CTGAGACAA	100
GagPol.ModSF	101	TGCTGTGGC	CAGCCCGAG	CTGGAGCGT	TCGCCGTGA	CCCGGGCTG	150
GagProt.ModS	101	TGCTGTGGC	CAGCCCGAG	CTGGAGCGT	TCGCCGTGA	CCCGGGCTG	150
Gag.ModSF2	101	TGCTGTGGC	CAGCCCGAG	CTGGAGCGT	TCGCCGTGA	CCCGGGCTG	150
GagPol.ModSF	151	CTGGAGACCA	GCGAGGGCTG	CCGCCAGATC	CTGGGCCAG	TGCAGCCCG	200
GagProt.ModS	151	CTGGAGACCA	GCGAGGGCTG	CCGCCAGATC	CTGGGCCAG	TGCAGCCCG	200
Gag.ModSF2	151	CTGGAGACCA	GCGAGGGCTG	CCGCCAGATC	CTGGGCCAG	TGCAGCCCG	200
GagPol.ModSF	201	CCTGCAGACC	GGCAGCGAG	AGCTCCGCG	CCTGTACAAC	ACCGTGCCCA	250
GagProt.ModS	201	CCTGCAGACC	GGCAGCGAG	AGCTCCGCG	CCTGTACAAC	ACCGTGCCCA	250
Gag.ModSF2	201	CCTGCAGACC	GGCAGCGAG	AGCTCCGCG	CCTGTACAAC	ACCGTGCCCA	250
GagPol.ModSF	251	CCTGTACTG	GCTGCACAC	CGCATCGAC	TCAAGGACAC	CAAGGAGGCC	300
GagProt.ModS	251	CCTGTACTG	GCTGCACAC	CGCATCGAC	TCAAGGACAC	CAAGGAGGCC	300
Gag.ModSF2	251	CCTGTACTG	GCTGCACAC	CGCATCGAC	TCAAGGACAC	CAAGGAGGCC	300
GagPol.ModSF	301	CTGGAGAGA	TGGAGGAGA	GCGAACACG	TCCAGAGAA	AGGCCACGCA	350
GagProt.ModS	301	CTGGAGAGA	TGGAGGAGA	GCGAACACG	TCCAGAGAA	AGGCCACGCA	350
Gag.ModSF2	301	CTGGAGAGA	TGGAGGAGA	GCGAACACG	TCCAGAGAA	AGGCCACGCA	350
GagPol.ModSF	351	GGCCGCGCC	GGCCGCGCA	CCGCCAGAC	CAGCCAGGTG	AGCCAGAACT	400
GagProt.ModS	351	GGCCGCGCC	GGCCGCGCA	CCGCCAGAC	CAGCCAGGTG	AGCCAGAACT	400
Gag.ModSF2	351	GGCCGCGCC	GGCCGCGCA	CCGCCAGAC	CAGCCAGGTG	AGCCAGAACT	400
GagPol.ModSF	401	ACCCATCGT	CGAGAACCTG	CAGGGCCAGA	TGGTGCACCA	GGCCATCAGC	450
GagProt.ModS	401	ACCCATCGT	CGAGAACCTG	CAGGGCCAGA	TGGTGCACCA	GGCCATCAGC	450
Gag.ModSF2	401	ACCCATCGT	CGAGAACCTG	CAGGGCCAGA	TGGTGCACCA	GGCCATCAGC	450

7 / 131

FIG. 7B

GagPol.ModSF	451	CCCCGACCC	470	480	490	500
GagProt.ModS	451	CCCGACACC	TGACGCTG	GTGAGAGTG	AGGCCTTCAG	500
Gag.ModSF2	451	CCCGACACC	TGACGCTG	GTGAGAGTG	AGGCCTTCAG	500
GagPol.ModSF	501	CCCGAGGAG	ATCCCATGT	TACGCGCCT	GACGAGGGC	550
GagProt.ModS	501	CCCGAGGAG	ATCCCATGT	TACGCGCCT	GACGAGGGC	550
Gag.ModSF2	501	CCCGAGGAG	ATCCCATGT	TACGCGCCT	GACGAGGGC	550
GagPol.ModSF	551	AGGACCTGAA	CACGATGTT	AACACCTGG	GCGGCCACCA	600
GagProt.ModS	551	AGGACCTGAA	CACGATGTT	AACACCTGG	GCGGCCACCA	600
Gag.ModSF2	551	AGGACCTGAA	CACGATGTT	AACACCTGG	GCGGCCACCA	600
GagPol.ModSF	601	CAGATCTGA	AGGAGACCAT	CAACGAGGAG	GCGCCGAGT	650
GagProt.ModS	601	CAGATCTGA	AGGAGACCAT	CAACGAGGAG	GCGCCGAGT	650
Gag.ModSF2	601	CAGATCTGA	AGGAGACCAT	CAACGAGGAG	GCGCCGAGT	650
GagPol.ModSF	651	GCACCCCGTG	CACGCCGGCC	CCATCGCCCC	CGCCGAGATG	700
GagProt.ModS	651	GCACCCCGTG	CACGCCGGCC	CCATCGCCCC	CGCCGAGATG	700
Gag.ModSF2	651	GCACCCCGTG	CACGCCGGCC	CCATCGCCCC	CGCCGAGATG	700
GagPol.ModSF	701	GGGCGACGCA	CATCGCGGC	ACCACACGCA	CCCTGACGGA	750
GagProt.ModS	701	GGGCGACGCA	CATCGCGGC	ACCACACGCA	CCCTGACGGA	750
Gag.ModSF2	701	GGGCGACGCA	CATCGCGGC	ACCACACGCA	CCCTGACGGA	750
GagPol.ModSF	751	TGGATGACCA	ACAACCCCC	CATCCCGTG	GGCGAGATCT	800
GagProt.ModS	751	TGGATGACCA	ACAACCCCC	CATCCCGTG	GGCGAGATCT	800
Gag.ModSF2	751	TGGATGACCA	ACAACCCCC	CATCCCGTG	GGCGAGATCT	800
GagPol.ModSF	801	GATCATCTTG	GGCTGACCA	AGATGCTGGG	GATGTACAGC	850
GagProt.ModS	801	GATCATCTTG	GGCTGACCA	AGATGCTGGG	GATGTACAGC	850
Gag.ModSF2	801	GATCATCTTG	GGCTGACCA	AGATGCTGGG	GATGTACAGC	850
GagPol.ModSF	851	TCTTGGACAT	CCCGCAGGGC	CCCAAGGAGC	CTTTCGGGAC	900
GagProt.ModS	851	TCTTGGACAT	CCCGCAGGGC	CCCAAGGAGC	CTTTCGGGAC	900
Gag.ModSF2	851	TCTTGGACAT	CCCGCAGGGC	CCCAAGGAGC	CTTTCGGGAC	900

8 / 131

GagPol..ModSF	901	CGCTTCTACA	AGACCCCTGCG	CGCTGAGCAG	GCACGACGAG	ACGTGAAGRA	950
GagProt..ModS	901	CGCTTCTACA	AGACCCCTGCG	CGCTGAGCAG	GCACGACGAG	ACGTGAAGRA	950
Gag..ModSF2	901	CGCTTCTACA	AGACCCCTGCG	CGCTGAGCAG	GCACGACGAG	ACGTGAAGRA	950
GagPol..ModSF	951	CTGGATGACG	GAGACCTTGC	TGGTGACAGAA	CGCCAAACCC	GACTGCAAGA	1000
GagProt..ModS	951	CTGGATGACG	GAGACCTTGC	TGGTGACAGAA	CGCCAAACCC	GACTGCAAGA	1000
Gag..ModSF2	951	CTGGATGACG	GAGACCTTGC	TGGTGACAGAA	CGCCAAACCC	GACTGCAAGA	1000
GagPol..ModSF	1001	CAATCTTGERA	GGCTCTGGCG	CCCGCGGCCA	CCCTGGAGGA	GATGATGACC	1050
GagProt..ModS	1001	CAATCTTGERA	GGCTCTGGCG	CCCGCGGCCA	CCCTGGAGGA	GATGATGACC	1050
Gag..ModSF2	1001	CAATCTTGERA	GGCTCTGGCG	CCCGCGGCCA	CCCTGGAGGA	GATGATGACC	1050
GagPol..ModSF	1051	GCCTGCGCAGG	GGTGGGCGG	CCCGCGGCCAC	AAGGCCCGCG	TGCTGGCCGA	1100
GagProt..ModS	1051	GCCTGCGCAGG	GGTGGGCGG	CCCGCGGCCAC	AAGGCCCGCG	TGCTGGCCGA	1100
Gag..ModSF2	1051	GCCTGCGCAGG	GGTGGGCGG	CCCGCGGCCAC	AAGGCCCGCG	TGCTGGCCGA	1100
GagPol..ModSF	1101	GGCGATGAGC	CAGGTGACGA	ACCCGCGGCAC	CATCATGATG	CAGCGCGGCA	1150
GagProt..ModS	1101	GGCGATGAGC	CAGGTGACGA	ACCCGCGGCAC	CATCATGATG	CAGCGCGGCA	1150
Gag..ModSF2	1101	GGCGATGAGC	CAGGTGACGA	ACCCGCGGCAC	CATCATGATG	CAGCGCGGCA	1150
GagPol..ModSF	1151	ACTTCCGCAA	CCAGCGGAG	ACCGTCAAGT	GCTTCACTG	CGGCNAGGAG	1200
GagProt..ModS	1151	ACTTCCGCAA	CCAGCGGAG	ACCGTCAAGT	GCTTCACTG	CGGCNAGGAG	1200
Gag..ModSF2	1151	ACTTCCGCAA	CCAGCGGAG	ACCGTCAAGT	GCTTCACTG	CGGCNAGGAG	1200
GagPol..ModSF	1201	GGCCACACCG	CCAGGAACCTG	CCGCGCCCGCC	CGCAAGAAGG	GCTGCTGGCG	1250
GagProt..ModS	1201	GGCCACACCG	CCAGGAACCTG	CCGCGCCCGCC	CGCAAGAAGG	GCTGCTGGCG	1250
Gag..ModSF2	1201	GGCCACACCG	CCAGGAACCTG	CCGCGCCCGCC	CGCAAGAAGG	GCTGCTGGCG	1250
GagPol..ModSF	1251	CTGCGGCGCCG	GNAGGACACC	AAATGAAGA	TTGACTGAG	AGACAGGCTA	1300
GagProt..ModS	1251	CTGCGGCGCCG	GNAGGACACC	AAATGAAGA	TTGACTGAG	AGACAGGCTA	1300
Gag..ModSF2	1251	CTGCGGCGCCG	GNAGGACACC	AAATGAAGA	TTGACTGAG	AGACAGGCTA	1300
GagPol..ModSF	1301	ATTTTITTAGG	GNAGATCTGG	CTTCTCTACA	AGGGAAGGCC	AGGGAATTTT	1350
GagProt..ModS	1301	ATTTTITTAGG	GNAGATCTGG	CTTCTCTACA	AGGGAAGGCC	AGGGAATTTT	1350
Gag..ModSF2	1301	ATTTTITTAGG	GNAGATCTGG	CTTCTCTACA	AGGGAAGGCC	AGGGAATTTT	1350

FIG. 7C

GagPol.ModSF	1351	1360	1370	1380	1390	1400
GagProt.Mods	1351	CTTCAGACGA	GACCAAGCC	RACAGCCCA	CCAGAGAGA	GCTTCAGGT
GagProt.Mods	1351	CTTCAGACGA	GACCAAGCC	RACAGCCCA	CCAGAGAGA	GCTTCAGGT
Gag.ModSF2	1351	CTCAGAGCC	GCCCGAGCC	CACGCCCCC	CCGAGAGGA	GCTTCGGTT
		1410	1420	1430	1440	1450
GagPol.ModSF	1401	TGGGGAGAG	AAACAATC	CCTCTAGAA	GCAGAGCGG	ATAGACAAGG
GagProt.Mods	1401	TGGGGAGAG	AAACAATC	CCTCTAGAA	GCAGAGCGG	ATAGACAAGG
GagProt.Mods	1401	TGGGGAGAG	AAACAATC	CCTCTAGAA	GCAGAGCGG	ATAGACAAGG
Gag.ModSF2	1401	CGCGAGAGG	AAAGCAACC	CCAGCCAGAA	GCAGAGCGG	ATAGACAAGG
		1460	1470	1480	1490	1500
GagPol.ModSF	1451	AACGTGATCC	TTTAACTTC	CTCAGATCA	TCITTTGGCAA	CGACCCCTCG
GagProt.Mods	1451	AACGTGATCC	TTTAACTTC	CTCAGATCA	TCITTTGGCAA	CGACCCCTCG
GagProt.Mods	1451	AACGTGATCC	TTTAACTTC	CTCAGATCA	TCITTTGGCAA	CGACCCCTCG
Gag.ModSF2	1451	AGCTGTATCC	CCTGACCAG	CTGCGAGCC	TGTTGGCAA	CGACCCCTCG
		1510	1520	1530	1540	1550
GagPol.ModSF	1501	TCACAGTAG	GATCGGGCC	CAGCTAAGG	AGCGCTGCT	CGACACGGC
GagProt.Mods	1501	TCACAGTAG	GATCGGGCC	CAGCTAAGG	AGCGCTGCT	CGACACGGC
GagProt.Mods	1501	TCACAGTAG	GATCGGGCC	CAGCTAAGG	AGCGCTGCT	CGACACGGC
Gag.ModSF2	1501	AGCCAGTAA
		1560	1570	1580	1590	1600
GagPol.ModSF	1551	GCCGACGCA	CCGTGCTGA	GGAGATGAAC	CTGCCCGGCA	AGTGAAGCC
GagProt.Mods	1551	GCCGACGCA	CCGTGCTGA	GGAGATGAAC	CTGCCCGGCA	AGTGAAGCC
GagProt.Mods	1551	GCCGACGCA	CCGTGCTGA	GGAGATGAAC	CTGCCCGGCA	AGTGAAGCC
Gag.ModSF2	1551
		1610	1620	1630	1640	1650
GagPol.ModSF	1601	CAAGATGATC	GCCGGATCG	GGGGCTTCAT	CAAGGTGGG	CAGTACGCC
GagProt.Mods	1601	CAAGATGATC	GCCGGATCG	GGGGCTTCAT	CAAGGTGGG	CAGTACGCC
GagProt.Mods	1601	CAAGATGATC	GCCGGATCG	GGGGCTTCAT	CAAGGTGGG	CAGTACGCC
Gag.ModSF2	1601
		1660	1670	1680	1690	1700
GagPol.ModSF	1651	AGATCCCGT	GGAGATCTGC	GGCCACAAGG	CCATCGGCAC	CGTCTGGTG
GagProt.Mods	1651	AGATCCCGT	GGAGATCTGC	GGCCACAAGG	CCATCGGCAC	CGTCTGGTG
GagProt.Mods	1651	AGATCCCGT	GGAGATCTGC	GGCCACAAGG	CCATCGGCAC	CGTCTGGTG
Gag.ModSF2	1651
		1710	1720	1730	1740	1750
GagPol.ModSF	1701	GGCCCCACC	CCGTGAACAT	CATCGGCCG	AACTGCTGA	CCAGATCGG
GagProt.Mods	1701	GGCCCCACC	CCGTGAACAT	CATCGGCCG	AACTGCTGA	CCAGATCGG
GagProt.Mods	1701	GGCCCCACC	CCGTGAACAT	CATCGGCCG	AACTGCTGA	CCAGATCGG
Gag.ModSF2	1701
		1760	1770	1780	1790	1800
GagPol.ModSF	1751	CTGACCCCTG	AACTTCCCA	TCAGCCCCAT	CGAGCGGTG	CCCGTGAAGC
GagProt.Mods	1751	CTGACCCCTG	AACTTCCCA	TCAGCCCCAT	CGAGCGGTG	CCCGTGAAGC
GagProt.Mods	1751	CTGACCCCTG	AACTTCCCA	TCAGCCCCAT	CGAGCGGTG	CCCGTGAAGC
Gag.ModSF2	1751

FIG. 7D

10/131

GagPol.ModSF 1801	1810	1820	1830	1840	1850
GagProt.ModS 1801	TGAAGCCGGG	GATGAGGG	CCCAAGGTCA	AGCAGTGGCC	CCTGACGAG
Gag.ModSF2 1801	TGAAGCCGGG	GATGAGGG	CCCAAGGTCA	AGCAGTGGCC	CCTGTAA...
GagPol.ModSF 1851	1860	1870	1880	1890	1900
GagProt.ModS 1851	GAGAGATCA	AGGCCCTGGT	GGAGATCTGC	ACCGAGATGG	AGAAGAGGG
Gag.ModSF2 1851	1910	1920	1930	1940	1950
GagPol.ModSF 1901	CAAGATCAGC	AAGATCGGCC	CCGAGAACCC	CTACAAACCC	CCCGTGTTCG
GagProt.ModS 1901	1960	1970	1980	1990	2000
Gag.ModSF2 1901	CCATCAAGAA	GAAGGACAGC	ACCAAGTGGC	GCAAGCTGGT	GGAATTCCGC
GagPol.ModSF 2001	2010	2020	2030	2040	2050
GagProt.ModS 2001	GAGCTGAACA	AGGCACCCA	GGACTTCTGG	GAGGTGCAGC	TGGGCATCCC
Gag.ModSF2 2001	2060	2070	2080	2090	2100
GagPol.ModSF 2051	CCAGCCCGCC	GGCTGAGAGA	AGAGAGAGAG	CGTGACCGTG	CTTGACGCTGS
GagProt.ModS 2051	2110	2120	2130	2140	2150
Gag.ModSF2 2051	CGGACGCTTA	CTTCAGCGTG	CCCTGGACA	AGGACTTCGG	CAAGTACACC
GagPol.ModSF 2101	2160	2170	2180	2190	2200
GagProt.ModS 2101	GCCTTCACCA	TCCCGAGCAT	CACACAGAG	ACCCCGGCGA	TCCGCTACCA
Gag.ModSF2 2101	2210	2220	2230	2240	2250
GagPol.ModSF 2151	GTACAAAGTG	CTGCCCCAGG	GCTGGAAGGG	CAGCCCCGCC	ATCTTCAGA
GagProt.ModS 2151	2260	2270	2280	2290	2300
Gag.ModSF2 2151	GTACAAAGTG	CTGCCCCAGG	GCTGGAAGGG	CAGCCCCGCC	ATCTTCAGA
GagPol.ModSF 2201	2310	2320	2330	2340	2350
GagProt.ModS 2201	GTACAAAGTG	CTGCCCCAGG	GCTGGAAGGG	CAGCCCCGCC	ATCTTCAGA
Gag.ModSF2 2201	2360	2370	2380	2390	2400
GagPol.ModSF 2251	GTACAAAGTG	CTGCCCCAGG	GCTGGAAGGG	CAGCCCCGCC	ATCTTCAGA
GagProt.ModS 2251	2410	2420	2430	2440	2450
Gag.ModSF2 2251	GTACAAAGTG	CTGCCCCAGG	GCTGGAAGGG	CAGCCCCGCC	ATCTTCAGA

FIG. 7E

GagPol.ModSF 2251	2260	2270	2280	2290	2300
GagProt.ModS 2251	GCAGCATGAC	CAAGATCTCG	GAGCCCTTCC	GCAAGCAGAA	CCCGGACATC
Gag.ModSF2 2251
	2310	2320	2330	2340	2350
GagPol.ModSF 2301	GTGATCTACC	AGTACATGGA	CGACCTGTAC	GTGGGACCG	ACCTGGAGAT
GagProt.ModS 2301
Gag.ModSF2 2301
	2360	2370	2380	2390	2400
GagPol.ModSF 2351	CGGCCAGCAC	CGCACCAAGA	TCGAGGAGCT	GCGCCAGCAC	CTGCTGGCT
GagProt.ModS 2351
Gag.ModSF2 2351
	2410	2420	2430	2440	2450
GagPol.ModSF 2401	GGGGCTTCAC	CACCCCGGAC	AGAGAGCACC	AGRAGAGGCC	CCCCCTTCCTG
GagProt.ModS 2401
Gag.ModSF2 2401
	2460	2470	2480	2490	2500
GagPol.ModSF 2451	TGGATGGGCT	ACGAGCTGCA	CCCCGACAAG	TGGACCGTGC	AGCCATCAT
GagProt.ModS 2451
Gag.ModSF2 2451
	2510	2520	2530	2540	2550
GagPol.ModSF 2501	GCTGCCCGAG	AGGACACGCT	GGACCGTGAA	CGACATCCAG	AAGCTGGTGG
GagProt.ModS 2501
Gag.ModSF2 2501
	2560	2570	2580	2590	2600
GagPol.ModSF 2551	GCAAGCTGAA	CTGGGCCAGC	CAGATCTACG	CCGGCATCAA	GGTGAGCAG
GagProt.ModS 2551
Gag.ModSF2 2551
	2610	2620	2630	2640	2650
GagPol.ModSF 2601	CTGTGAAGC	TGCTGGCGGG	CACCAAGGCC	CTGACCGAGG	TGATCCCCCT
GagProt.ModS 2601
Gag.ModSF2 2601
	2660	2670	2680	2690	2700
GagPol.ModSF 2651	GACCGAGGAG	GCCGAGCTGG	AGCTGGCCGA	GAACCGCGAG	ATCCTGAAGG
GagProt.ModS 2651
Gag.ModSF2 2651

FIG. 7F

12 / 131

GagPol.ModSF 2701	2710	2720	2730	2740	2750
GagProt.Mods 2701	AGCCCGTCA	CGAGGTGTAC	TAGCAGCCCA	GCAAGGACCT	GTGTGCCGAG
Gag.ModSF2 2701
GagPol.ModSF 2751	2760	2770	2780	2790	2800
GagProt.Mods 2751	ATCCAGAAGC	AGGCCAGGG	CCAGTGGACC	TACCAGATCT	ACCAGGAGCC
Gag.ModSF2 2751
GagPol.ModSF 2801	2810	2820	2830	2840	2850
GagProt.Mods 2801	CTTCAAGAAC	CTGAAGACCG	GCAAGTAGCC	CCGATGCGC	GGCGCCACA
Gag.ModSF2 2801
GagPol.ModSF 2851	2860	2870	2880	2890	2900
GagProt.Mods 2851	CCAACGCGT	GAAGCAGCTG	ACCGAGGCCG	TGCAGAGGT	GAGCAGCCAG
Gag.ModSF2 2851
GagPol.ModSF 2901	2910	2920	2930	2940	2950
GagProt.Mods 2901	AGCATCGTGA	TCTGGGCAA	GATCCCAAG	TTCAGCTGC	CCATCCAGAA
Gag.ModSF2 2901
GagPol.ModSF 2951	2960	2970	2980	2990	3000
GagProt.Mods 2951	GGAGACCTGG	GAGGCTTGGT	GGATGGAGTA	CTGGCAGGCC	ACCTGGATCC
Gag.ModSF2 2951
GagPol.ModSF 3001	3010	3020	3030	3040	3050
GagProt.Mods 3001	CCGAGTGGGA	GTTCTGTAAC	ACCCGCCCCG	TGCTGAAGCT	GTGGTACCAG
Gag.ModSF2 3001
GagPol.ModSF 3051	3060	3070	3080	3090	3100
GagProt.Mods 3051	CTGGAGAGGG	AGCCCATCGT	GGGCGCCGAG	ACCTTCTACG	TGGACGGCGC
Gag.ModSF2 3051
GagPol.ModSF 3101	3110	3120	3130	3140	3150
GagProt.Mods 3101	CCGCAACCCG	GAGACCAAGC	TGGGCGAGCG	CGGCTACGTG	ACCGAGCCGCG
Gag.ModSF2 3101

FIG. 7G

GagPol.ModSF 3151	3160	3170	3180	3190	3200
GagProt.Mods 3151	GCCTCCAGAA	GGTGTGTGAC	ATCGCCGACA	CCACCAACCA	GAAGACCGAG
Gag.ModSF2 3151
GagPol.ModSF 3201	3210	3220	3230	3240	3250
GagProt.Mods 3201	CTGCNAGCCA	TCACCTGGC	CTGCAGGAC	AGCGGCTTGG	AGGTGACAT
Gag.ModSF2 3201
GagPol.ModSF 3251	3260	3270	3280	3290	3300
GagProt.Mods 3251	CGTGACCGAC	AGCCAGTAGC	CCTGGGCAT	CATCCAGGCC	CAGCCCGACA
Gag.ModSF2 3251
GagPol.ModSF 3301	3310	3320	3330	3340	3350
GagProt.Mods 3301	AGAGCGAGAG	CGAGCTGGTG	AGCCAGATCA	TCGAGCAGCT	GATCAAGAAG
Gag.ModSF2 3301
GagPol.ModSF 3351	3360	3370	3380	3390	3400
GagProt.Mods 3351	GAGAAGGTGT	ACCTGCGCTG	GGTCCCGGCC	CACAAGGCCA	TCGGCGGCAA
Gag.ModSF2 3351
GagPol.ModSF 3401	3410	3420	3430	3440	3450
GagProt.Mods 3401	CGAGCAGGTG	GACAACTGG	TGAGCGCGG	CATCCGCAAG	GTGCTGTTC
Gag.ModSF2 3401
GagPol.ModSF 3451	3460	3470	3480	3490	3500
GagProt.Mods 3451	TGAAATGGCAT	CGTAAAGGC	CAGGAGGAGC	ACGAGAAGTA	CCACAGCAC
Gag.ModSF2 3451
GagPol.ModSF 3501	3510	3520	3530	3540	3550
GagProt.Mods 3501	TGGCCGCGCA	TGGCCAGCA	CTTCAACCTG	CCCTCCGCTGG	TGGCCAGGGA
Gag.ModSF2 3501
GagPol.ModSF 3551	3560	3570	3580	3590	3600
GagProt.Mods 3551	GATCTGTGCC	AGCTGCGACA	AGTGCCAGCT	GAAGGGCGAG	GCCATGCAG
Gag.ModSF2 3551

FIG. 7H

GagPol.ModSF	3601	3610	3620	3630	3640	3650
GagProt.ModS	3601	CTCGAGTGG	GGCATCTGGC	AGTGTGCTG	CACCCACCTG	3650
Gag.ModSF2	3601	3650
GagPol.ModSF	3651	3660	3670	3680	3690	3700
GagProt.ModS	3651	GAGGCGAAGA	TGATCTGTGT	GGCGGTGCAC	GTGCGCCAGC	GCTACATCGA
Gag.ModSF2	3651	3700
GagPol.ModSF	3701	3710	3720	3730	3740	3750
GagProt.ModS	3701	GGCCGAGGTG	ATCCGCGCG	AGACCGGCCA	GGAGACCGCC	TACTTCTGCG
Gag.ModSF2	3701	3750
GagPol.ModSF	3751	3760	3770	3780	3790	3800
GagProt.ModS	3751	TGAAGCTGGC	CGGCCGCTGG	CCCTGTGAAGA	CCATCCACAC	CGACAGCGGC
Gag.ModSF2	3751	3800
GagPol.ModSF	3801	3810	3820	3830	3840	3850
GagProt.ModS	3801	AGCNACTTCA	CCAGCACCCAC	CGTGAAGGCC	GCCTGTCTGT	GGCGCGGCAT
Gag.ModSF2	3801	3850
GagPol.ModSF	3851	3860	3870	3880	3890	3900
GagProt.ModS	3851	CAAGCAGGAG	TTGCGCATCC	CCTACAACCC	CCAGAGCCAG	GGCGTGTGG
Gag.ModSF2	3851	3900
GagPol.ModSF	3901	3910	3920	3930	3940	3950
GagProt.ModS	3901	AGAGCATGAA	CAAGCAGCTG	AAGAAGATCA	TGCGCCAGGT	GCAGACCCAG
Gag.ModSF2	3901	3950
GagPol.ModSF	3951	3960	3970	3980	3990	4000
GagProt.ModS	3951	GCCGAGCACC	TGAAGACCGC	CGTGCAGATG	GGCGTGTTC	TCCACACTT
Gag.ModSF2	3951	4000
GagPol.ModSF	4001	4010	4020	4030	4040	4050
GagProt.ModS	4001	CAAGCGCAAG	GCGCGCATCG	GCGGTACAG	CGCCGCGCAG	CGCATCTGG
Gag.ModSF2	4001	4050

FIG. 71

15 / 131

GagPol.ModSF	4051	4060	4070	4080	4090	4100
GagProt.ModS	4051	ACATCATCGC	CACCGACATC	CAGACCAAGG	AGCTGCAGAA	GCAGATCACC
Gag.ModSF2	4051
GagPol.ModSF	4101	4110	4120	4130	4140	4150
GagProt.ModS	4101	AGATCCAGAA	ACTTCGGCGT	GTACTACCGC	GACACACAGG	ACCCCTGTGC
Gag.ModSF2	4101
GagPol.ModSF	4151	4160	4170	4180	4190	4200
GagProt.ModS	4151	GAAGGCGCC	GCCAACTGCG	TGTGGAAGGG	CGAGGGCGCC	GTGTGTATCC
Gag.ModSF2	4151
GagPol.ModSF	4201	4210	4220	4230	4240	4250
GagProt.ModS	4201	AGGACACAG	CGACATCAAG	GTGTGTGCCC	GCCGCAAGGC	CAAGATCATC
Gag.ModSF2	4201
GagPol.ModSF	4251	4260	4270	4280	4290	4300
GagProt.ModS	4251	CGCGACTACG	GCAAGCAGAT	GCGCGGCGAC	GACTTGGTGG	CCAGCCGCCA
Gag.ModSF2	4251
GagPol.ModSF	4301	4310	4320	4330	4340	4350
GagProt.ModS	4301	GGACGAGGAC	TAG.....
Gag.ModSF2	4301

FIG. 7J

16 / 131

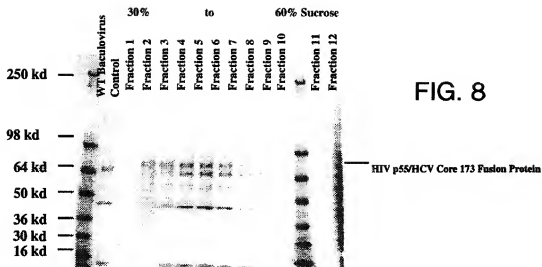


FIG. 8

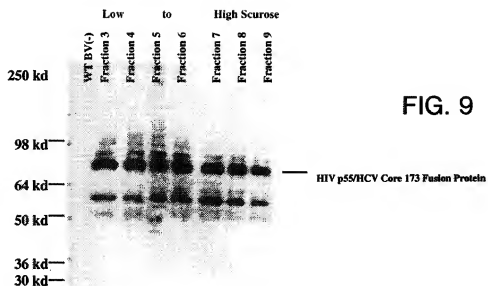


FIG. 9

17 / 131

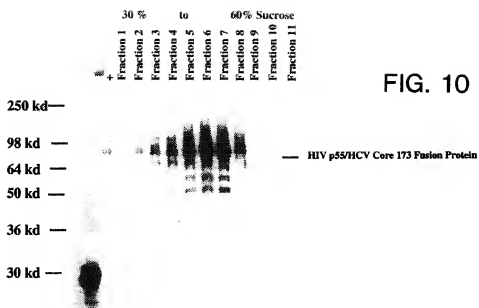


FIG. 10

18 / 131

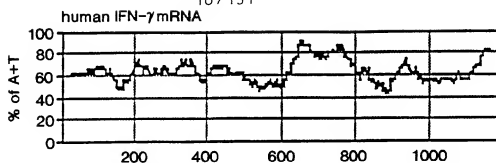


FIG. 11A

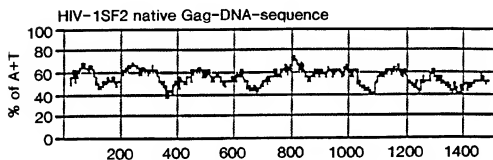


FIG. 11B

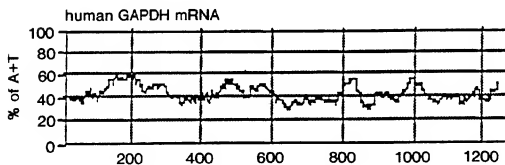


FIG. 11C

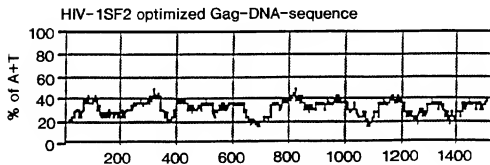


FIG. 11D

19 / 131

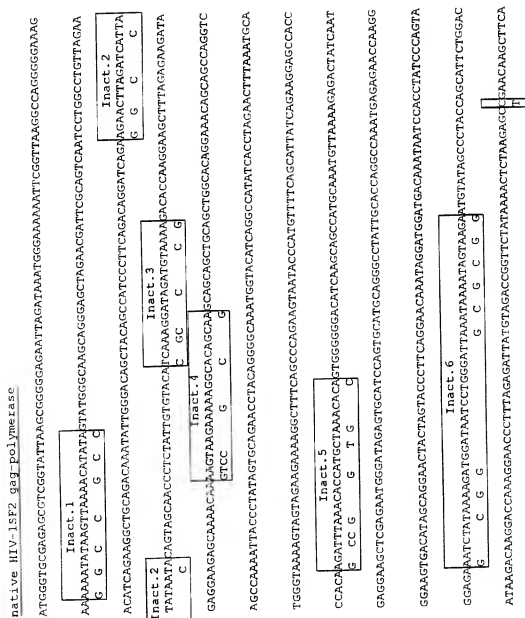


FIG. 12A

20/131

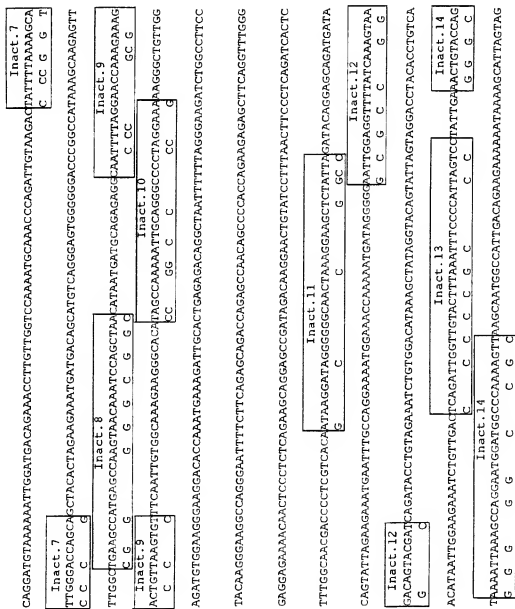


FIG. 12B

FIG. 12C

AGATATGTACAGAAATGGAAAGGAGGAAATTTCAAAAATTTGGGCTGGAATCCATCAATACTCCAGTATTG
 CTATAAGAAAAAGACAGTACTAAATGGAGAAAACTAGTAGATTTACAGAGACCTTAATAAAGAACTCAGACCTCT
 GGGAGATTCAGTTAGGAATACACACCCCGAGGGTTAAAAAGAAATAACAGTACAGATTTGGATGTGGGTGATG
 CATATCTTTTCAGTTTAAAGACTTTAAAGACTTACTGCATTAACACTAGTATATAACAACTAGACAC
 CAGGATTAAGATACATCAATGTGCTCCAGGATGGAAGATCACCAAGCAATTTCCAAAGTAGCATGACAA
 AAATCTTAGAGCCTTTTAGAAACAGAAATCCACACATGTTATCTATCAATACATGGATGATTTGTATGTAGGATCTG
 ACTTAAGATAGGCGCATAGACAAAAATAGGAACTGACACAGCATCTGTGAGGTGGGATTTACCAACCAG
 ACRAAACAATCAGAAAGACCTCCATCTCTTTGGTGGCTTATGAATCCATCTGATAAATGGACAGTACAGCTTA
 TTTATGCCAGGGATTAAGTTAAGCGATTTGTAAATCTCTTAGGGAACCAAGCACTACAGATTAATACCACTAA
 CAAAGAGACCTAGTAGCAAAATACAGACAGGGCAGGCCAATGGACATATCAAAATTTACAGAGGCACTTAA
 ATCTGAAAACAGAAAGTATGCAAGGATGAGGGTGCCACACTAATGATGTAAACAGTTTACAGAGGCGAGTGCAAA
 AAGTATCCACAGAAAGCATAGTAATATGGGAAAGATTCCTAAATTTAAACTACCCATACAAAGGAAACATGGGAAG
 CATGGTGGTGAAGTATTTGGCAAGCTACTCGGATTCCTGAGTGGGAGTTTGTCAATACCCCTCCCTTAGTGAATAT
 GTTACAGTTTAGAGAAAGACCCCATAGTAGGAGCAAGAACTTTCTATGTAGTGGGCGAGCTAATAGGAGACTAAAT
 TAGGAACAGCAGGNTATCTTACTGCAGAGGACAAAAGTTGTCTCCATAGCTGACACACAAATCAGAGACTG
 AATTACAGCAGNTATCTTACTTTGCAGGATTCGGGATTAGAATTAACATAGTAAACAGCTCAACACTCAATATGCATTAG
 GAATCAITTCAGCAACACAGATAGAGTGAAATCAGATAGTCTAGTCAAAATAGAGAGGATTAATTAAGAGGAA
 AGTCTCACTGGCATGGGTACCAAGCACAAAGGAAATGGAGAAATGAACAGTAGATTAATATTAGTCTGCTGGAA
 TCAGAAAGTACTAATTTTGAATGGAAATAGGAGCCCAAGAGAACATGAGAAATATCACAGTAATTTGGAGGCAA
 TGCTAGTGTATTTAACTTGCCACTGTAGTACAAAGAAATAGTACCGAGCTGTGAATATCTAGTAAAGGAG
 AAGCCATGCATGGACAGTAGTGTAGTCCAGGAATGGCACTAGATTGTACACATCTAGAAGGAAATATCC
 TGTAGCATTTCAATGTAGCAGTGTGATATAGAGCAGAAAGTTATCCAGCAGACAGGCGAGGAAACAGCATATT
 TTCTCTTAAATTAGCAGGAAGTGGCCAGTAAACAAATACATACAGCAATGSCAGCAATTTCCAGTACTACGG
 TTATGGCCGCTGTGGTGGGCGAGTCAAGCAAGAAATTTGGCAATCCCTACATCCCAAAAGTCAAGGAGTAGTAG
 TAATCTATGAATTAATGAATTAAGAAATTTAGGACAGGTAAGAGATCAGGCTGAACACTTAAGACAGCATACAA
 TGGCAGATATCTCCAAATTTAAAGAAAGGGGGATTTGGGGATACAGTGCAGGGGAAAGATAGTAGACATA
 TACCAAGATCCCTTTGGAAAGGACACAGAAAGCTCTCTGGAAGTGTGAAGGGCGAGTAGTAAACAGAAATAGTGTG
 ACATAAAGTAGTGTGCCAGAGAAAGGCAAAATCTATTAGGGATTTGMAAAACAGATGGCAGGTTGATTTGTGTGG
 CAAATAGACAGGATGAGGNTATG

22 / 131

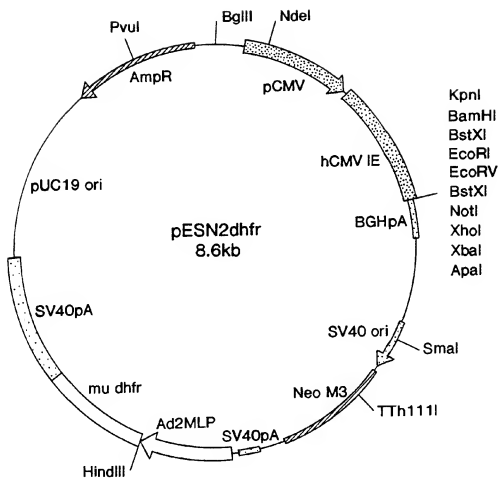


FIG. 13A

23 / 131

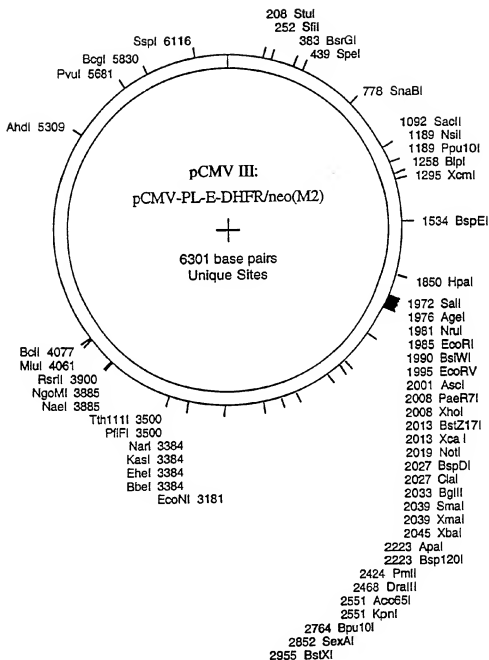


FIG. 13B

24 / 131

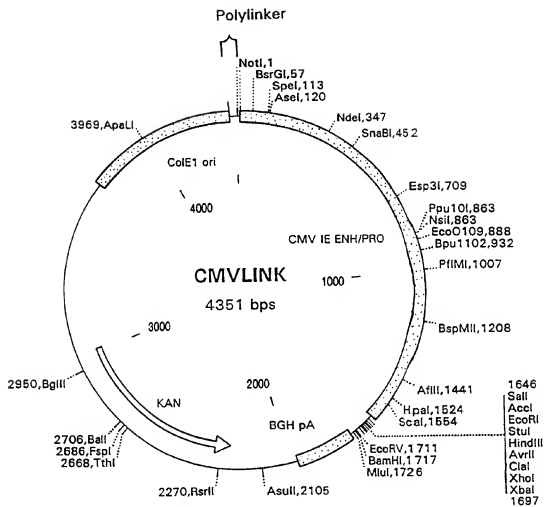


FIG. 14

25 / 131

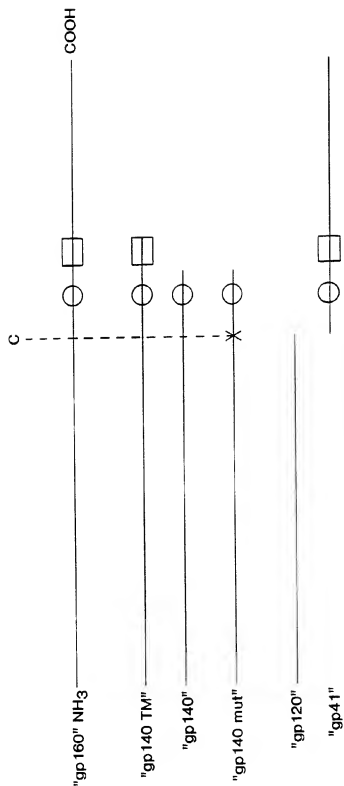


FIG. 15

gp120wtSF162

GTAGAAAATTGGGTGCAGTCTATTATGGGTACTGTGTGGAAGAAGCAACACCACCTCTATTTT
GTGCAATCAGATGCTTAAGCTATGACACAGAGGTACATAATGTCTGGGCCACACATGCTGTGTACCCAC
AGACCTTAACCCACAGAAATAGTATTGGAAAATGTGACAGAAAATTTAACTGTGGGAAAATAACRATG
GTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAAGTCTAAAGCCATGTGTAAGTTAAACCC
CACTCTGTGTTACTCTACATTGCATTAATTGGAAGATGCTACTAATCCNAGAGTAGTAATTGGAAAGA
GATGACACAGAGGAGAAATAAAAAATTGCTCTTTCAAGGTCACCAAGCATTAAGAAAATTAAGATGCAGAAA
GAATATGCACTTTTATAAATGATAGTACCAATAGATTAATGATAATACAAGCTATAAAATTGATAA
ATTGTAACACCTCAGTCATACAGGCCTGTCCAAAGGTATCCTTTGAAACCAATCCCATACATATTG
TGCCCGGCTGGTTTTGGCATCTAAAGTGTAAATGATAAGAAGTTCAATGGATCAGGACCATGTACAAT
GTCACACAGTAGACAATGTACACATGGAATTAGCCAGTAGTGTCAACTCAATTCCTGTTAAATGGCAGTC
TAGCAGAAAGAGGGGTAGTAATTAGATCTGAAATTTTCACAGACAATGCTTAAACTATAATAGTACAGCT
GAAGGAATCTGTAGAAAATTAAATGTACAAGACCTTAACAATAACAAGAAAAGTATAACTATAGGACCG
GGGAGAGCATTTATGCAACAGGAGCATATAGGAGATATAAGACAAGCACAATTGTAACATTAGTGGAG
AAAAATGGAAATAACCTTTAAAAACAGATGTACAAAATTACAAGCAATTTGGGAATAAACAATAGT
CTTTAAGCAATCTCAGAGGGGACCCAGAAAATTGTAATGCACAGTTTAAATTTGGAGGGGAATTTTTC
TACTGTAATTCACACAGCTTTTAAATAGTACTTGGAAATAACTATAGGGCCAAATAACACTAATGGAA
CTATCACCTCCCATGCAGAAATAAACAATTTAAACAAGGTGGCAGGAAGTGAAGAAAAGCAATGTATGC
CCCTCCCATCAGAGGACAAATTAGATGCTCATCAAAATTTACAGGACTGCTATTAAACAAGAGATGGTGGT
AAAGAGATCAGTRACACCCAGAGATCTTCAGACCTTGGAGTGGAGATATGNGGACAAATTGGAGAGTIG
AATTATATAATATAAGTAGTAAAAAATTGAGCCATTAGGAGTAGCACCCACCAAGGCAAGAGAGAGT
GGTGCAGAGAGNAAAAAGA

FIG. 16
(SEQ ID NO:30)

gp140wtSF162

GTAGAAAAATTTGGGTCACAGTCTATTATGKGGTACCTGTGTGGAAGAACACACCTCCTCTATTTT
 GTGCATCAGATGTAAAGCCTATGACACAGAGGTACATAATGTCTGGGCCACACATGCCCTGTGTAACCAAC
 AGACCTTACCCCAAGAAATAGTATTTGGAAATGTGACAGAAAAATTTTAACTGTGGAAAAATTAACATG
 GTAGACAGATGTCATGGGATATAATCAGTTTATGGGATCAAGTCTAAAGCCATGTGTAAAGTTTAAACCC
 CACTGTGTACTCTACATTTGACATTAATTTGAAGAATGCTACTAATACCAGAGTAGTAAATTTGGAAGA
 GATGGACAGAGGAAAAATAAAATTTGCTTTTCAAGGTACACACAGCATAAAGAAATAAGATGCGAAGA
 GAAATGCACTTTTATTAACTGTAGTACCAATAGATATGATAATACRAGCTATAAATTGATATAA
 AATTGAACCTCAGTCAITACACAGGCTGTCCAAAGGTATCTTTTGAACCAATTTCCCATACATTAATTG
 TGCCTCGGCTGTGTTTTCGATTCCTAAAGTGTAAATGATAAGAAAGTTCAATGGATCAGGACCATGTACAAT
 GTCAGCACATACAAATGTACACATGGAAATTAGGCCAGTAGTGCAACTCAATTCGTGTTAAATGGCAGTC
 TAGCAGAAGAGGGTATGTAATTTAGTACTGMAAATTTTCCAGACATGCTTAAACTATAATAGTACAGCT
 GAAGGAATCTGTAGAAATTAATTTGTAACAAGACCTAACTAAATAATACAGAAATAAGTATACTATAGGACCG
 GGGAGAGCATTTTATGCAACAGGAGACATAATAGGAGATATAAGACAGCACATTTGTAACATTAGTGGAG
 AAAAAATGGAATACCTTTAAACAGAGATAGTTACAAAATTAACAGCAATTTGGGAATTAACAACTAATG
 CTTTAAAGCAATCTCAGGAGGGGACCCAGAAATTTGTAATGCACACTTTTAAATTTGGAGGGAAATTTTC
 TACTGTAAATTCACACAGCTTTTAAATAGTACTTGGAAATACTATAGGGCCAAATAACACTAATGGAA
 CTAACACTCCCAATGAGAAATAAACAAATTTATAACAGGTGGCAGGAAGTAGGAAGCAATGTATGC
 CCCTCCCATCAGAGGACAAATTTAGATGCTCATCAATATTACAGGACTGCTATTAAACAGAGATGCTGTT
 AAAGATCTAGTAAACACCCAGATCTTCAGACTGGAGGTGGAGATATGAGGACAAATTTGGAGAGTG
 AATTATAATAATATAAGTAGTAAATTTGAGCCATTTAGGATAGTACCCACCAAGGCAAGAGAGAGT
 GGTGCAGAGAGAAAAAGAGCAGTACGCTAGGAGCTATGTTCTTGGTCTCTGGGAGCAGCAGGAAGC
 ACTATGGGGCAACGCTCACTACGCTAGGTAACGGCCAGCAATTTGTTGGTATAGTGCACAGC
 AGCAAAATTTGCTGAGAGCTATTGAGCGCCACACAGCTCTGTTGCACTCAGCTCTGGGGCATCAAGCA
 GCTCCAGGAGAGTGTCTGCTGTGGGAAGATACCTTAAAGGATCAACAGCTCTTAGGGATTTTGGGGTGC
 TCTGGAAATCTATTGCAACCACTCTGCTGCTGGGAATGCTAGTTGGAGTAATAAATCTCTGGATCAGA
 TTTTGGAAATAACATGACCTGAGATGGAGTGGGAGAGAGAAATGCAAAATTAACAACTTAATATACACTT
 AATTGGAATTCGCAAGACCAACAGAAAGAAATGAACAAGAAATTTAGAAATGGATAGTGGGCAAGT
 TTTGGGAATTTGGTTTGGACATATCAAAATGGCTGTGGTATATA

FIG. 17
 (SEQ ID NO.31)

28 / 131

gp160wtSF162

GTAGAAAAATTTGGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAAGAAGCAACCACCCTCTATTTT
GTGCATCAGATGCTAAAGCCTATGACACAGAGGTACATATGCTCGGGCCACACATGCCTGTGTACCCAC
AGACCCCTAACCCACAAGAAATAGTATTGGAAAAATGTGACAGAAAAATTTTAACATGTGGAAAAATAACATG
GTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAAGTCTAAAGCCATGTGTAAAGTTAAACCC
CACTCTGTGTTACTCTACATTGCCTAAATTTGAAGAAATGCTACTAATACCAAGAGTAGTAATTTGGAAAAA
GATGGACAGAGGAGAAAAATAAAAAATTTGCTCTTTCAAGGTCAACCAAGCATAGAATAAGATGCAGAAA
GAATATGCACCTTTTTATAAATCTTGATGTAGTACCAATAGATAATGATAATACAAGCTATAAATTTGATAA
ATTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAACCAATTCCTATACATTATTG
TGCCCCGGCTGGTTTGGCATTCTAAAGTGTAAATGATAAGAAAGTTCAATGGATCAGGACCATGTACAAAT
GTCAGCACAGTACAAATGTACACATGGAATTAGGCCAGTAGTGTCAACTCAATTGCTGTTAAATGGCAGTC
TAGCAGAAGAAGGGGTAGTAATTAGATCTGAAAAATTTACAGACAATGCTAAAACTATAATAGTACAGCT
GAAGGAATCTGTAGAAATTAATTTGACAAGACCTAACAAATAATACAGAAAAAGTATAACTATAGGACCG
GGGAGAGCATTTTATGCAACAGGAGACATAATAGGAGATATAAGACAAGCACATTTGGAATAAAACAATAGT
AAAAATGGAATAACACTTTAAAAACAGATAGTTACAAAATTAACAGCACAAATTTGGGAATAAAACAATAGT
CTTTAAGCAATCTCAGGAGGGGACCCAGAAAAATGTAATGTCACAGTTTAAATTTGGAGGGGAATTTTTTC
TACTGTAATTTCAACACAGCTTTTTAATAGTACTTGGAAATAATCATATAGGGCCAAATTAACACTAATGGAA
CTATCACACTCCCATGCGAATAAAAACAAATTATAAACAGGTGGCAGGAAATAGGAAAAAGCAATGTATGC
CCCTCCCATCAGAGGACAAATTAGATGCTCATCAAAATATTACAGGACTGCTATTAAACAAGAGATGGTGGT
AAAGAGATCAGTAACACCACCAGATCTTCAGACCTGGAGGTGGAGATATGAGGGAACAATTTGGAGAAATG
AATTATATAAATAAAGTAGTAAAAATTTAGGCCATTAGGAGTAGCACCACCAAGGCAAGAGAGAAAGT
GGTGACAGAGAAAAAAGAGCAGTGACGCTAGGAGCTATGTTCTTGGGTCTTTGGGAGCAGCAGGAAGC
ACTATGGGCGCACGGTCACTGACGCTGACGGTACAGGCCAGACAAATTTGCTGGTATAGTGCAACAGC
AGAACAATTTGCTGAGAGCTATTGAGGCGCAACAGCATCTGTTGCAACTACAGCTCTGGGGCATCAAGCA
GCTCCAGGCAAGAGTCTCGCTGTGGAAGATACCTAAAGGATCAACAGCTCCTAGGGATTTGGGGTTGC
TCTGAAAACTCATTTGCCACCCTGCTGTGCTTTGGAATGCTAGTTGGAGTAATAAATCTCTGGATCAGA
TTTGGAAATACATGACCTGGATGGAGTGGGAGAGAGAAAAATGACAATTACACAAACTTAATATACACCTT
AATTGAAGAAATCGCAGAACCAACAGAAAGAAATGAAACAAAGATATTAGAATTGATAGTGGGCAAGT
TTGTGGAATTTGGTTTGACATATCAAAATGGCTGTGGTATATAAAAAATATTCAATATAGTAGTAGGAGGT
TAGTAGGTTAAGGATAGTTTTTACTGTGCTTTCTATAGTGAATAGAGTTAGGCAGGGATCTCACCATT
ATCATTTTCAGACCCGCTTTCCAGCCCCAAGGGGACCCGACAGGCCGGAAGGAATCGAAAGAAAGGTTGGA
GAGAGAGACAGAGACAGATCCAGTCCATTAGTGCAATGGATTATTAGCACTCATCTGGGACGATCTACGGA
GCCTGTGCCTCTTCAGTACCACCCTGTGAGAGCTTAATCTTGATTGTCAGCGAGGATTGTGGAACTTCT
GGGACGACGGGGTGGGAAGCCCTCAAGTATTGGGGGAATCTCTGCAGTATTGGATTACAGGAATTAAG
AATAGTCTGTAGTTGTTTGTATGCCATAGCTATAGCAGTAGCTGAGGGGACAGATAGGATTATAGAAG
TAGCAAAAGAAATTTGGTAGAGCTTTTCTCCACATACCTAGAAGAAATAAGCAGGGCTTTGAAAGGGCTTT
GCTATAA

FIG. 18

(SEQ ID NO:32)

gp120.modSF162

gaattcgccaccatggatgcaatgaagagggtctgctgtgtgctgctgtgtgtggagcagtc
 ttcgtttcgccagcgctgagaagctgtgggtgacgtgtactacggcgtgccgtgtggag
 gaggccaccaccctgtctgcccagcagcccaaggctacgacaccggagtgcacaaagtg
 tggccaccacgcgtggtgccaccgaccccaaccacaggagatcgtgtggagaacgtgacc
 gagaactcaactgtggaagaacaactggtggagcagatgcacgaggacatcatcagcctgtgg
 gaccagacctgaagcctggtgaagctgacccctgtggtgacctgactgcaccaactg
 aagaacgcccacacccaagcagcaactggaaggagatggaccggcgagatcaagaactgc
 agcttcaagtgaccaccagcatccgcaacagatgcagaaggagtacgaccttctacaagctg
 gacgtggtgccatcgacaacgacacacacagctacaagctgatcaactgcacaccagcgtgatc
 accaggcctgccccaggtagcttcyagccccatccccatccactactgcgcccccgcggttc
 gccatcctgaagtgcaacgacaagaagtcaacggcagcgccctgacccaactgagcacctg
 cagtgcaaccaaggcatcgccccgtggtgagcaccagctgctgctgaacggcagcctggccgag
 gagggcgtggtgatccgcagcgaacttcccgcaacgccaagaccatcatcgtgcagctgaag
 gagagcgtggagatcaactgcaccccccacaacaacacccgcaagagcatcaccatcgcccc
 ggccgccttctacgccaccggcgacatctcgcgacatccggcaggccactgcacatcagc
 ggcgagaagtggaaacaacacctgaagcagatcgtgaccaagctgcaggccagcttcggcaacag
 accatcgtgttcaagcagagcagcgggcagccccgagatcgtgatgcacagcttcaactgcggc
 ggcagttcttctactgcacagcaccagctgttcaacagacctggaacacacccatcgcccc
 aacaaaccaaaggcaccatcacctgccctggccatcaagcagatcatcaaccgtggcaggag
 gtgggcaaggccatgtacgccccccatccggccagatccgctgcagcagcaacatcacggc
 ctgctgctgacccggacggcggaaggagatcagcaacacacccagagatcttcgcccccgggc
 ggccagatgcgcgcaactggcgcgagcagctgtcaagtacaaggtggtgaagtcgagccccg
 ggcggtggccccccaccaggccaaagcgccgctggtgcagcgcgagaagcgctaacctcgag

FIG. 19
(SEQ ID NO:33)

gp120.modsF162.delV2

gaattcgccaccatgatgcaatgaagagagggtctctgctgtgctgctgctgtgtggagcagtc
 ttctgttgcgccagccgtggagaagctgtgggtgacctgtactacggcggtgcccggtgtggaag
 gaggccaccaccacctgttctgcccagcagcccaaggcctacgacaccgaggtgcacaacgtg
 tgggccaacccacgctgcgtgccaccgaccccaacccaccaggagatcgtgctggagaaactgacc
 ggaacttcaacatgtggaagaacaatgltggagcagatgcacaggacatcatcagctgtgg
 gaccagacctgaagcctgctgaagctgacccctgtgctgacctgacctgacctgacccaactg
 aagaaagcccaacacacagagcagcaactggaaggagatggaccggcgagatcaagaactgc
 agcttcaaggtggcgccggcaagctgatcaactgcaacaccagctgatcaccagggcctgccc
 aaggtgagcttcgagcccatcccatccactgcgccccgcggcttcgcatcctgaagtgc
 aacgacaagaagtcaacggcagggccctgcaccaactgagcaccgtcagtgacccacggc
 atccgcccgtggtgagcaccagctgctgtgaaaggcagctggccgaggaaggcgtggtgatc
 cgagcgagaactcaaccgacaacgccaagacctcatcgtgcagctgaagagagcgtggagatc
 aactgcaacccgccaacaacaacccgcaagagcatcaccatcgccccggcgccctctac
 gccaccggcagatcatcggcgacatccgcccaggccactgcaacatcagcgcgagaagtggaaac
 aacacctgaagcagatcgtgaccaagctgcaggccagttcggaacaagacctcgtgtcaag
 cagagcagcgcgccgaccccgagatcgtgatgcacagcttcaactggcgcgagttcttctac
 tgcacaacgacccagctgttcaacagcaccctgaaacaacacctcgccccaaacaacaccaggc
 accatcacctgccttgcgcatcaagcagatcatcaacgctggcagaggtgggcaaggccatg
 tacgcccccccatccgcgccagatccgtgcagcagcaatcacccgctgctgacccgc
 gagggcgcaaggagatcagcaacaccaccagatcttcgccccggcgcgagatgcgcgac
 aactggcgagcgagctgtacaagtacaaggtggtgaagtcgagccctggcggtggccccacc
 aaggccaagcgcgctggtgcagcgcgagaagcgtcaactcgag

FIG. 20
 (SEQ ID NO:34)

gp120.modSF162.delV1V2

gaatcgcaccatggatgcaatgaagagagggtctctgctgtgtgctgctgtgtgagcagtc
 tctgttctgccagccgtgagaagctgtgggtgacgtgtactacggcgtgccgtgtggaag
 gaggccaccaccctgtctcgcgcagcagcccaaggcctacgacacggagtgcaacaagtg
 tggccacccacgcctgcgtgccaccgaccccaacccacaggagatcgtgctggagaagctgacc
 gaaactcaacatgtgaagaacaacatggtggagcagatgcacggagacatcatcagcctgtgg
 gaccagacctgaagccctgcgtgaagctgacccctgtgcgtggcgccggaactgccagacc
 agcgtgatcccccaggcctgcccaaggtgagcttcgagcccatcccatccactactgcgcctcc
 gccggttcgcatcctgaagtgaacgacaagaagtccaacggcagcggccctgcaccaactg
 agcacctgcaagtcaaccacggcatccgccctgggtgagcaaccagctgctgtgaacggcagc
 ctggccgagagggtggtggtgatccgagcgagaactcaccgacaacgcccaagaccatcatcgtg
 cagctgaaggagagcgtggagatcaactgcaaccgcccccaacaacaacccgcaagagcatcacc
 atcggcccccgcgcctctcagccaccggcagatcatcggcgacatccggcaggccactgc
 aacatcagcggcgaagtggaaacacacctgaagcagatcgtgaccaagctgaaggccagttc
 ggcaacaagaccatcgtgtcaagcagcagcggcgaccctcccgagatcgtgatgcacagcttc
 aactgcggcggcagttcttctactgaacagcaccacagctgtcaacagcactggaaacaacc
 atcggccccaacaacaccacggcaccatcacctgcctgcgcgatcaagcagatcatcaaccgc
 tggcaggaggtgggcaagccatgtacgcccccccatccggccagatccgctgcagcagcaac
 atcaaccgctgctgctgacccgcagcggcgcaaggagatcagcaacaccaccagatcttcgcg
 ccggcgccgcgcacatgcgacaactggcgcagcagctgtacaagtacaagggtgggtgaagtc
 gaggcctggcggtggcccccaaccaggccagcgcgtggtgcagcgcgagagagcgtcaact
 gag

FIG. 21
(SEQ ID NO:35)

32 / 131

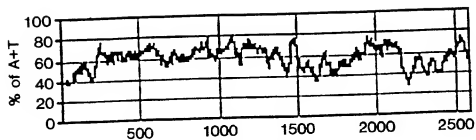


FIG. 22A

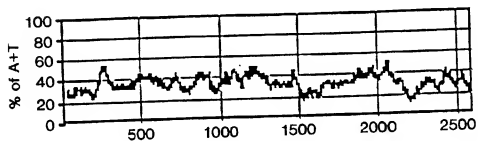


FIG. 22B

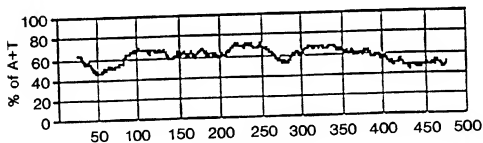


FIG. 22C

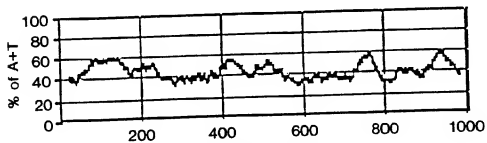


FIG. 22D

33 / 131

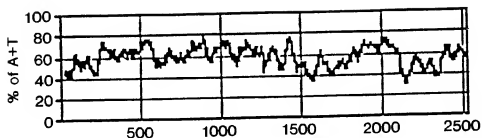


FIG. 22E

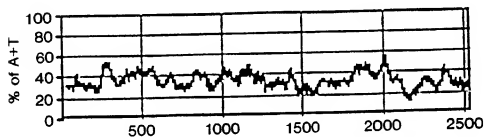


FIG. 22F

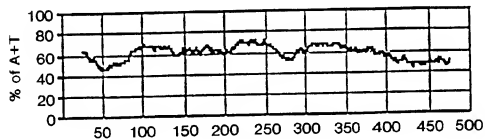


FIG. 22G

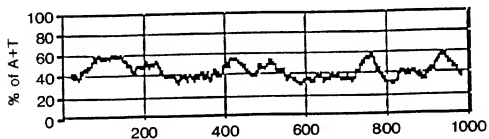


FIG. 22H

gp140.modSF162

gaattcgccaccatggatgcaatgaagagagggctctgctggtgtgctgctgctggtgtggagcagtc
 ttcggtttcgccagcgccgtggagaagctgtgggtgacccgtgtactacggcgctgcccgtgtggaaag
 gagggccaccaaccctgttctgcccagcgacgccaaggcctacgacaccgaggtgcacaacgtg
 tgggcccaccacgctctgctgcccacgaccccaacccccaggagatcgtgctggagaacgtgacc
 gagaacttcaacatgtggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg
 gaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgcaactgcaaccaacctg
 aagaaagccaccacccaagagcagcaactggaaaggagatggacccgaggcagatcaagaactgc
 agcttcaagggtgaccaccagcatccgcaacaagatgcagaaggagtagcggcctgttctacaagctg
 gacgtggtgcccattcgacaacgcaacaccagctacaagctgatcaactgcaacaccagcgtgatc
 acccaggcctgccccagggtgagcttcgagcccatccccatccaactactgccccccgcccggcttc
 gccatcctgaagtgcacgacaagaagtgcacggcagcgggccctgcaaccaacgtgagcaccctg
 cagtgacccacggcatccgccccgtggtgagcaccacagctgctgctgaacggcagcctggccgag
 gagggcgtggtgatccgagcgagaaacttcaccgacaacgccaagaccatcatcgtgagcgtgaag
 gagagcgtggagatcaactgcaccgcccccaacaacaaccccgcaagagcatcaccatcgcccc
 ggccgccccttctacgcccagcgacatcatcgggacatcccgccaggcccatgcaacatcagc
 ggcgagaagtggaaacaacaccctgaagcagatcgtgaccaagctgcaggcccatgcccgaacaag
 accatcgtgtcaagcgagcagcgccggcgaccccgagatcgtgatgcacagcttcaactcgccg
 ggagagttcttctactgcacaacagcaccagctgttcaacagcacctggaaacaacaccatcgcccc
 aacaacaccaacggccaccatcacccctgcctgcccgcataagagcagatcatcaaccgctggcaggag
 gtgggcaaggccatgtacgcccccccatccgcccggcagatccgctgcagcagcaacatcacgggc
 ctgctgctgaccccgacggcgccgaaggagatcagcaaacaccacagagatcttccgccccggcgggc
 ggcgacatgcgcgacaactggcgccagcgagctgtacaagttacaaggctggtgaagatcgagccctg
 ggctggcccccccaaggccaagcgccgctgggtgcagcgcgagaagcgccgctgacccctgggc
 gccatgttccctgggcttctggggcgccgcccggcagcaccatggggcccccagcctgacccctgacc
 gtgcaggcccgcagctgctgagcggcatcgtgcagcagcagaacaacctgctgcccgcacatcgag
 gccagcagcacctgctgcagctgacccgtgtggggcatcaagcagctgcaggcccgctgctggcc
 gtggagcgctacctgaaggaccagcagctgctggggcatctggggctgcagcggcgaagctgatctgc
 accaccgcccgtgcccctggaaagccagctggagcaacaagagcctggaccagatctggaaacaacatg
 acctggatggagtgggagcgcgagatcgacaactacaccaacctgatctacacctgatcgaggag
 agccagaaaccagcaggagaagaacagcagcaggagctgctggagctggacaagctggggccagcctgtgg
 aactggttcgacatcagcaagtggctgtggtacatctcaactcgag

FIG. 23
(SEQ ID NO:36)

35/131

gpi140.modSF162.delV2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgtgtgtggagcagtc
 ttgctttcgccacgcgcgtggagaagctgtgggtgacccgtgtactacggcgtgcccgtgtggaa
 gaggccaccaccacctgttctgcgccagcgagccaggcctacgacaccgaggtgcacaaactgt
 tggggccacccacgcctgctgcccaccgaccccaacccccaggagatcgtgctggagaaactgtgac
 gagaacttcaacatgtggaaagaacaacatgggtggagcagatgcacaggagacatcatcagcctgtgg
 gaccagagcctgaaagcctgctgctgaagctgacccccctgtgctgacccctgcaactgcaccaaactg
 aagaacccacccaacacccaagagcagcaactggaaaggagatgggacccggcgagatcaagaactgc
 agcttcaaggtggggcgccggcaagctgatcaactgcacaccagcgtgatcaccaggcctgcccc
 aaggtgagcttcgagcccatccccatccactactgcgcccccgccggcttcgccatcctgaagtgc
 aacgacaagaagtccaacggcagcggccccctgcaccaactgtgagcaccctgcagtgacccaacggc
 atccgccccgtgggtgagcaccagctgctgctgaacggcagcctggccgaggaggggcgtggtgatc
 cgcagcgagaaacttcacgcacaacgcccaagacatcatcgtgcagctgaaggagagcgtggagatc
 aactgcacccgcgcccaacaacaccccgcaagagcatcaccatcgcccccgccgcgccttcctac
 gccacccggcgacatcatcgggcgacatccgcccaggcccactgcaacatcagccggcgagaagtggaa
 aacacccctgaagcagatcgtgaccaagctgcaggcccaacttcggcaacaagacatcgtgtcaag
 cagagcagcggcgccgaccccgagatcgtgatgcacagcttcaactgcggcgagtgcttctctac
 tgcacacgacccacagctgttcaacagcaccctggaaacaacacatcggccccaacaacaccaaaggc
 accatcacccctgcctgcgcgatcaagcagatcatcaacgcctggcaggaggtgggcaaggccatg
 tagccccccccatccggggccagatccgctgcagcagcaacatcacccggcctgctgctgacccgc
 gacggcggaaggagatcagcaacaccccgagatcttcgccccggcgccggcgacatgcgcgac
 aactggcgcgagcgtgtacaagtacaagtggtgaagatcgagccccctggcgctggccccccacc
 aaggccaagcgccgctgtgtgctgcagcgcgagaagcgccgctgacccctggcgccatgttctggggc
 ttcctggcgccggcgccgagcaccatggggcgcccgagcctgacccctgacgtgcagggcccgccag
 ctgctgagcggcctcgtgcagcagcagaacaacccctgctgcggccatcgaggccagcagcaccctg
 ctgcagctgacccgtgtggggcatcaagcagctgcaggcccgccgtgtgctgctggagcgtacctg
 aaggacccagcagctgctgggcatctggggctgcagcggcaagctgatctgcacccaccgctgccc
 tggaaacgcagctggagcaacaagagcctggaccagatctggaaacaacatgacctggatggagtgg
 gagcgcgagatcgacaactacccaacctgatctacacccctgatcgaggagagccgaacacagcag
 gagaagaacgagcaggagctgctggagctggacaagtggggcagccctgtggaaactgggtcgacatc
 agcaagtggctgtggatcatctaacctcgag

FIG. 24

(SEQ ID NO:37)

36 / 131

gp140.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgtgtgtggagcagtc
ttcgtttcgccccagcgccgtggagaagctgtgggtgacgctgtactacggcgtgcccgtgtggaa
gagggccaccaccacctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaaactg
tgggccaaccacgcctgctgctgccaccgacccccaccccgaggatcgtgctggagaactgacc
gagaacttcaacatgtggagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgacccccctgtgctgtgggcccggcaactgccagacc
agcgtgatcaccaggccctgccccagggtgagcttcgagcccatccccatccactactgcccctcc
gccggtctcgccatcctgaagtgcacgacaagaagttcaacggcagcgccctgcaccaactgtg
agcacccgtgcagtgcacccacggcatccgccccgtggtgagcaccagctgctgctgaacggcagc
ctggcccgagggggctgggtgatccgcagcgagaacttcaaccgacaacccaagaccatcatcgtg
cagctgaaggagagcgtggagatcaactgcacccgccccacaacaacacccgcgaagagcatcacc
atcgcccccggccgcgccttctacgccaccggcgacatcatcggcgacatccgcccaggccactgc
aacatcagcggcgagaagtggaaacaacacctgaagcagatcgtgcaccaagctgcaggccctgc
ggcaacaagaccatcgtgttcaagcagagcagcgccggcgacccccagatcgtgatgcacagcttc
aactgcccggcgaggttcctactgcaacagcaccacagctgttcaacagcacctggaaacaacac
atcgcccccacaacaccaacggccaccatcacctgccctgcccgcatacagcagatcatcaaccgc
tggcaggaggtgggcaaggccatgtacgccccccccatccgcccagatccgctgcagcagcaac
atcacccggctgctgctgaccgcgacggcgccgaaggagatcagcaacacacccagatctccgc
cccggcgcgcgacatgcgcgacaactggcgagcgagctgtacaagtaacaaggtggtagagatc
gagccccctggcgctggccccaccgaaggccaagcgccgcgtggtagcgcgagagaagcgcgcgtg
acctggggcgccatgttcctgggcttcctgggcgcgcgcggcagcaccatggggcccccgcagcctg
acctgacccgtgcaggcccgccagctgctgagcggcatcgtgcagcagcagaacaacctgctgcgc
gccatcgaggccagcagcacctgctgcagctgacctgtggggcatcaagcagctgcaggccgc
gtcgtggccgtggagcgtacctgaaggaccagcagctgctgggcatctggggctgcagcggcaag
ctgatctgcaccacgcgcgtgcccgggaacgcgcagctggagcaacaagagcctggaccagatctgg
aacaacatgacctggatggagtgaggcgcgagatcgacaactacaccaacctgatcacacctg
atcgaggagagccagaaccagcaggagaagaacgagcagagctgcaggagctggaagctggagcc
agcctgtggaaactggctcgacatcagcaagctggctgtggtagatctaacctcgag

FIG. 25
(SEQ ID NO:38)

37 / 131

gp140.mut.modSF162

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtctt
 cgtttcgccagcgccgtggagaagctgtgggtgaccgtgtactacggcgctgcccgtgtggaa
 gaggccaccaccacctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcaacaactgt
 tggggccaccacgcctgctgctgcccacgaccccaacccccaggagatcgtgctggagaactgtgac
 gagaactcaacatgtggaagaacaacatgggtggagcagatgacagaggacatcatcagcctgtgg
 gaccagagcctgaagccctgctggaagctgacccccctgtgctgtgacctgacactgacccaacctg
 aagaacgccaccaacacagagcagcaactgggaaggagatggaccggcgagatcaagaactgc
 agctcaaggtgaccaccagcatccgcaacaagatgcagaaggagtagccctgttctacaagctg
 gacgtgggtgccccatcgacaacgacaacaccagctacaagctgatcaactgcaacaccagcgtgatc
 acccaggcctgccccagggtgagcttcgagcccatccccatccactactgccccccgcccgttcc
 gccatcctgaagtgcacgacaagaagtcaacggcagcggccccctgacccaactgagcaccgtg
 cagtgcacccacggcatccgccccgtgggtgagcaccagctgctgtgtaacggcagcctggccgag
 gagggcgtggtgatccgagcagagaacttcaccgacaacgcccaagaccatcatcgtgcagctgaag
 gagagcgtggagatcaactgcaccgcccccaacaacaacaccgcgaagagcatcaccatcgcccc
 ggccgcgcccttctacgcccacggcgacatcatcgccgacatccgcccaggccccactgcaacatcagc
 ggcgagaagtgcacaacacacctgaagcagatcgtgaccaaagctgcaggccccagtccggcaacaag
 accatcgtgttcaagcagagcagcggcgccgaccccgagatcgtgatgcacagcttcaactcgccg
 ggcgagtcttcttactgcaacagcaccacagctgttcaacagcaccctggacaacaacccatcgcccc
 aacaacaccaacggccacatcaccctgcccgtgcgcgatcaagcagatcatcaaccgtggcaggag
 gtgggcaaggccatgtacgccccccccatccgcccagatccgctgcagcagcaacatcacccg
 ctgctgctgacccgcgacggcgccgaaggagatcagcaacaccaccgagatcttccgccccggcgccg
 ggcgacatgcccgaacaactggcgagcagcgtgtacaagtaacaagtggtgaagatcgagccccctg
 ggctggcccccccaaggccaagcgcccgctggtgcagcgcgagaagagcgccgtgacctggggc
 gccatgttccctgggttccctggggcgccgcccgcagcaccatggggcgcccgagcctgacctgacc
 gtgcaggcccgccagctgctgagcggcatcgtgcagcagcagacaacctgctgcgcgccatcgag
 gccagcagcaccctgctgcagctgacccgtgtggggcatcaagcagctgcaggcccgcgctgtggcc
 gtggagcgctacctgaaggaccagcagctgctgggcatctggggcgccgaagagcctggaccagatctggaacaacatg
 accacggccgctgcccctggaaacggccagctggagcacaacagcctggaccagatctggaacaacatg
 acctggatggagtgggagcgcgagatcgacaactacacacacacctgatacaccctgatcgaggag
 agccagaacacagcaggagaagaacagagcaggagctgctggagctggacaagtggggcccagcctgtgg
 aactggttcgacatcagcaagtggctgtgggtacatctcaactcgag

FIG. 26

(SEQ ID NO.39)

gp140.mut.modSF162.delV2

gaattcgccaccatggatgcaatgaagagaggggctctgctgtgtgctgctgctgtgtggagcagt
tctggtttcgccagcgccgtggagaagctgtgggtgacgctgtactacggcgctgcccgctgtggaag
gaggccaccaccctgttctgcgcccagcgacgccaaaggcctacgacaccgaggtgcacaacgtg
tgggcccacccacgctgtgctgcccacgaccccaacccccaggagatcgtgctgggaacagtgaac
gagaacttcaacatgtggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgacccccctgtgcgtgacccctgcactgcaccaacctg
aagaacgccaccaaccaaagcagcaactgggaaggagatggaccgcggcgagatcaagaactgc
agcttcaaggtggggcgccggcaagctgatcaactgcaacaccagcgtgatcaccaggcctgcccc
aagggtgagcttcgagcccatcccatccactactgcgcccccgccgcttcgccatcctgaagtgc
aacgacaagaagtccaacggcagcggccccctgcaccaacgtgagcaccgctgcagtgcacccacggc
atccgccccctgtggtgagcaccacgtgctgctgaaaggcagcctggccgaggaggggcggtggtgatc
cgagcgagaaacttcaccgacaacgccaagacatcatcgtgcagctgaaggagagcgtggagatc
aactgcaccgcgcccccaacaacacccgcaagagcatcaccatcgggccccggccgcgcttctac
gccaccggcgacatcatcggcgacatccgcccaggccactgcaacatcagcggcgagaaagtggaa
aacaccctgaagcagatcgtgaccaagctgcaggccacgttcggcaacaagacatcgtgtcaag
cagagcagcggcgggcgaccccgagatcgtgatgcacagcttcaactgcggcgggcgagttcttctac
tgcaacagcaccacagctgttcaacagcacctggaaacaacacatcggcccccaacaacacacggc
accatcaccttgcctcgccgatcaagcagatcatcaaccgctggcaggaggtgggcaaggccatg
tacgcccccccatccgcgccagatccgctgcagcagcaacatcaccggcctgcctgctgaccgcg
gacggcgggcaaggagatcagcaacaccacagagatcttccgccccggggcgggcgacatgcgcgac
aactggcgcgagcgtgtacaagtaacaagtggtgaagatcgagccctggggcgtggccccccac
aaggccaagcgccgctggtgcagcgcggaagagcgccgtgacccctggggcgccatgttctctggg
ttctctggggcgccgcccggcagcacatggggccccgcagccctgacccctgacccgtgcaggccccgcag
ctgctgagcgggcatcgtgcagcagcagaacaacctgctgcgcgccatcgaggcccagcagcactg
ctgcagctgacgctgtggggcatcaagcagctgcaggcccgctgctggcgctggagcgctacctg
aaggaccagcagctgctgggcatctggggctgcagcggcgaagctgatctgcaccaccccgctgccc
tggaacgccagctggagcaacaagagcctggaccagatctggaacaacatgacctggatggagtg
gagcgcgagatcgacaactacaccaacctgatctacacctgatcgaggagagccagaacccagcag
gagaagaacgagcaggagctgctggagctggacaagtgggccagcctgtggaaactggttcgacatc
agcaactggctgtggtacatctaaactcgag

FIG. 27

(SEQ ID NO:40)

39 / 131

gp140.mut.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc
 ttcgtttcgccagcgccgaggagaagctgtgggtgacccgtgactacggcgctgcccgtgtggaag
 gagggccaccaccacccctgttctgcgcgcagcgacgccaaaggccctacgacaccgaggtgcacaacgtg
 tgggcccaccacgcccgtgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc
 gagaacttcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg
 gaccagagcctgaagccctgcgtgaagctgacccccctgtgctggggcgccggcaactgcccagacc
 agcgtgatcaccagggcctgcccgaaggtgagcttcgagcccacccccatccactactgcgcccc
 gccggccttcgcccctcctgaagtgcacgcgaagaagttcaacggcagcgggccccctgcaccaacgtg
 agcaccgtgcagtgaccccacggcatccgcccctgtgtgagcaccacgtgctgctgaacggcgagc
 ctggccgaggagggcgtggatccgcagcagagaacttcaccgacacgcgccaagaccatcatcgtg
 cagctgaaggagagcgtggagatcaactgcaccgcgcccacaacaacacccgcgaagagcatcacc
 atcgcccccgccgcccctctacgcaccggcgacatcatcgggacatccgcagggccactgc
 aacatcagcggcgagaagtggaaacaacccctgaagcagatcgtgaccaagctgcaggcccagttc
 ggcaacaagaccatcgtgttcaagcagagcagcggcgagcccgagatcgtgatgcacagcttc
 aactcggcgccggaggtctctctactgcaacagcaccagctgttcaacagcacctggaaacaaccc
 atcgcccccaacaacacccagccaccatcacccctgccctgccgcacgaagcagatcatcaaccgc
 tggcaggaggtgggcaaggccatgtacgcccccccatccgcccagatccgctgcagcagcaac
 atcacccgcctgctgctgacccgcgacggcggaaggagatcagcaacaccccgagatcttcg
 cccggcgggcgagcatgcgcgaacaactggcgagcagcgtgtacaagtaacaaggtggtagaatc
 gagccccctggcggtggcccccccaaggccaagcgccgctggtgcagcgcgagaagagcgccgtg
 accctgggcgccatgttccctgggcttcctgggcgcgcccggcagcaccaaggcgcccgcagcctg
 accctgacccgtgcaggcccgcagctgctgagcggcatcgtgcagcagcaacaacctgctgcgc
 gccatcgaggccacagcacctgctgcagctgacccgtgtggggcatcaagcagctgcaggcccgc
 gtgctggccgtggagcgtacctgaaggaccagcagctgctgggcatctggggctgcagcggcaag
 ctgatctgcacccacggcgtgcccggaaacgcagctggagcaacaagagcctggaccagatctgg
 aacaacatgacctggatggagtgggagcgcgagatcgacaactacaccaacctgatctacacctg
 atcgaggagagccagaacagcaggagaagaacgagcagggagctgctggagctggacaagtgggac
 agcctgtggaactggttcgacatcagcaagtgcgtgtggtacatctaactcgag

FIG. 28
(SEQ ID NO:41)

gp140.mut7.modSF162

gaattcgccaccatggatgcaatgaagagaggggctctgctgtgtgctgctgctgtgtggagcagtctcgttctcgccacgcgccgtggagaagctgtgggtgacgctgtactacggcgctgcccgctgggaagaggcccaaccaacctggttctgcccagcgacgccaaggcctacgacaccgaggtgcaacaagtggtggccacccacgctgccccacgacgcccccaaccccccaggagatcgtgctggagaacgtgaccgagaaactcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatagccctgtgggaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgcaactgcaccaacctgaagaaacgccaaccaaccaagagcagcaactgggaaggagatggacccggcgagatcaagaactgcagcttcaaggtgacccaccagcatccgcaacaagatgcagaaggagtagcgcctgttctacaagctggacgtggtgccccatcgacaacgacaacacagctacaagctgatcaactgcacaacacagcgtgatcaccaggcctgccccaggtgagcttcgagcccatccccatccaactactgcgcccccgccggcttcgccatcctggaagtgcacaacgacaagaagtcaacggcgagggccccctgcaccaacgtgagcaaccgtgcagtgcaacccacggcatccgcccctggtgagcaccacagctgctgctgaaacggcgagcctggccgagaggggcgtggtgatccgacgaggaacttcaccgacaacgccaagaccatcatcgtgcagctgaaaggagcgtggagatcaactgcacccgcccccaacaacaccccgcaagagcatcaccatcgccccggccgcgccttctacgcccacggcgacatcatcggcgacatccgacaggcccaactgcaacatcagcggcgagaagtggaaacaacacctgaagcagatcgtgaccaagctgcaggccagcttcggcaacaagaacctcgtgtcaagcagagcagcggcgccgacccccgagatcgtgatgcacagcttcaactcgccggcgagcttcttctactgcgaacagcaccacagctgttcaacagcacctggaaacaacacatcgccccaaacaacaccaacggccacatcacctgcccctgcgcgatcaagcagatcatcaacccgtggcgaggaggtgggcaaggccatgtacgccccccccatccgcccgcagatccgctgcagcagcaacatcacccgctgctgctgaaccccgacggcgccgaaggagatcagcaacacacccagatcttccgccccggcgccggcgacatcgcgacaactggcgagcgagctgtacaagtacaagtggtgaaagatcgagccccgtggcgtggccccccaaggccatcagcagcgtggtgcagagcgagaagagcgccgtgaacctggggcgccatggtcctgggcttctggggcgcccgccgagcaccatggggcgcccgagcctgacctgaccgtgacgtgcaggcccccgagctgctgagcggcatcgtgcagcagcagaaacacctgctgcgcgccatcgagcccagcagcacctgctgcagctgacccgtgtggggcatcaagcagctgcaggcccccgctgctggccgtggagcgtcactgaaggacacgacgtgctgggcatctggggctgcagcggcaagctgatctgcaccaacccgcgtgcccggagcgcagatcgacaactacaccaacctgatctacacccctgatcgaggagagccagaaacagcaggagaagaaacgagcaggagctgctggagctggacaagcgggccagcctgtgggaactggttcgacatcagcaagtggctggtgtacatctaactcgag

FIG. 29

(SEQ ID NO:42)

gp140.mut7.modSF162.delV2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc
 ttctgtttcgcccgccgctggagaagctgtgggtgaccgtgtactacggcgctgcccgctgtggaaag
 gaggccaccaccacctgttctgcccagcgaagcgaagcctacgacaccgaggtgcaacaactgtg
 tgggcccacccacgctgctgtcccacgaccccaacccccaggagatcgtgctggagaacgtgacc
 gagaacttcaacatgtggaagaacaacatgggtggagcagatgcacaggacatcatcagcctgtgg
 gaccagagcctgaagccctgctggaagctgacccccctgtgctgacctgcaactgcaacaacctg
 aagaacgccaaccaaccaagagcagcaactggaaggagatggaccgctggagatcaagaactgc
 agctcaagggtgggcccggcaagctgatcaactgcaacaccagcgtgatcacccaggcctgcccc
 aagggtgagcttcgagcccatccccatccactactgcgcccccgccggcttcgccaatcctgaagtg
 aacgacaagaagtcaacggcagcggccccctgcacaaacgtgagcaccgtgcagtgcaaccacggc
 atccgcccctgtgtgagcaccagctgctgctgaacggcagcctggccgaggaggggcgtgggtgatc
 cgcagcgagaaacttaccgacaacgcgaagaccatcatcgtgcagctgaaggagagcgtggagatc
 aactgcaacccgccccacaacaacacccgcgaagacatcaccatcgccccggccgcttctac
 gccaccggcgacatcatcggcgacatccgcccaggccactgcaacatcagcggcgagaaagtggaa
 aacacctgaagcagatcgtgaccaagctgcaggccagcttcgggaacaagaccatcgtgtcaag
 cagagcagcggcgccgaccccgagatcgtgatgcacagcttcaactgcggcgagcttcttctac
 tgcaacagcaccagctgttcaacagcaccctgggaacaacaccatcgcccccaacaacacacggc
 accatcacccctgcccctgccgcatcaagcagatcatcaaccgctggcaggaggtggggcaaggcctg
 tacgccccccccatccgcccagatccgctgcagcagcaacatcacggcctgtgctgcaaccgc
 gaccgcccgaaggagatcagcaacacacacagagatcttccgccccggcgggcgagatgcgcgac
 aactggcgagcagcgtgtacaagtacaagtggtgaagatcgagccccctggcgctggccccccac
 aaggccatcagcagcgtggtgcagagcgaagaagcgcctgacccctggcgccatgttcttggg
 tctctggcgccgcccggcagcaccatgggcccggcagcctgacccctgacccgtgcaggcccgccg
 ctgctgagcggcatcgtgcagcagcagaacaacctgctgcgcccctcagggcccagcagcaccctg
 ctgcagctgacccgtgtggggcatcaagcagctgcaggcccgctgctggccgtggagcgttaacctg
 aaggaccagcagctgctgggcatctggggctgcagcggcgaagctgatctgcaccaccccgctgccc
 tggaaacgcagctggagcaacaagagcctggaccagatctggaacaacatgacctggatggagtg
 gagcgcagatcgacaactacaccaacctgatctacacccctgacgaggagagccgaacacagcag
 gagaagaaagcagcaggagctgctggagctggacaagctgggcccagcctgtggaactgggtcgacatc
 agcaagtgctgtggtacatctaactcgag

FIG. 30

(SEQ ID NO:43)

42 / 131

gpl40.mut7.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc
 ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgctgccctgttggaaag
 gagggccaccaccacctgtctcgccgagcgagcgaaggcctacgacacccgaggtgcacaacgtg
 tgggcccacccacgcctgctgcccacgcaccccaaccccccaggagatcgtgctggagaacgtgacc
 gagaacttcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg
 gaccagagcctgaagcctgctggaagctgacccccctgtgctggggcgccgggcaactgcccagacc
 agcgtgatcaccacggcctgcccgaagggtgagcttcgagccccatccccatccaactactgcgcccc
 gccggcttcgccaatcctgaagtgcacgcacaagaagtccaacggcagcgccccctgcaccaacgtg
 agcacctgtagtgacccccacggcatccgccccgtggtgagcaccacgtgctgctgaaacggcagc
 ctggccgaggaggcgctggtgatccgcagcgagaacttcacgcacaacgcgaagaccatcatcgtg
 cagctgaaggagagcgtggagatcaactgcacccgccccaaacaacaaccccgcaagagcatcacc
 atcggcccccgccgcgccttctacgcccacggcgacatcatcggcgacatccgcccaggcccactgc
 aacatcagcgggcgagaagtggaaacaacacctgaagcagatcgtgaccaagctgcaggcccagttc
 ggcaacaagaccatcgtgttcaagcagagcagcgggcggaaccccgagatcgtgatgcacagcttc
 aactgcggcgggcagttcttctactgcaacagcacccagctgttcaacagcacttggaaacaacac
 atcggcccccaacaacccaacggcaccatcacctgcctgcccgcagatcaagcagatcatcaacgcg
 tggcaggagggtgggcaaggccatgtacgccccccccatccgcccagatccgctgcagcagcaac
 atcacccggcctgctgtgacccgcgacggcgggcaaggagatcagcaacacacccagagatctccgc
 cccggcgggcgacatgcgcgacaactggcgagcgagctgtacaagtacaagggtggtgaagatc
 gagccccctggcggtggcccccaacaaaggccatcagcagcgtggtgcagagcgagaagagcgccgtg
 accctggggcgccatgttcctgggcttcctggggcgccggcgagcaccatggggccccgcagcctg
 accctgacctgcaggcccccgagctgctgagccgcatcgtgcagcagcagaacaacctgctgcgc
 gccatcgaggccacagcagcacctgctgcagctgacccgtgtggggcatcaagcagctgcaggccccg
 gtgctggccgtggagcgctacctgaaggacacagcagctgctggggcatctggggctgcagcggcgaag
 ctgatctgcacccacccgctgccccggaaacgcagctggagcaacaagagcctggaccagatctgg
 aacaacatgacctggatggagtgggagcgcgagatcgacaactacacaaacctgatctacacctg
 atcgaggagagccagaaacacagcaggagaagaacagcaggagctgctggagctggacaagctggggc
 agcctgtggaactggttcgacatcagcaagtggctggtacatctaaactcgag

FIG. 31

(SEQ ID NO:44)

44 / 131

gp140.mut8.modSF162.delV2

gaattcgccaccatggatgcaatgaagagaggggctctgctgtgtgctgctgctgtgtggagcagtc
 ttcgtttcgcccgagcgccgtggagaagctgtgggtgacgctgtactacggcgctgcccgctgtggaag
 gaggccaccaccacccctgttctgcccagcgacgccaaggccctacgacacccgaggtgcacaacgtg
 tggggccaccacgctgctgctgcccacccgaccccaaccccgaggagatcgctgtggagaacgtgacc
 gagaacttcaacatgtggagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg
 gaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgacatgcaccaacctg
 aagaacgccaccaacacccaagagcagcaactggaaggagatggaccgcccggcgagatcaagaactgc
 agcttcaagggtgggcccggcaagctgatcaactgcaacaccagcctgatcacccaggcctgcccc
 aagggtgagcttcgagcccatccccatccactactgcccccccgccgcttcgccatcctggaagtgc
 aacgacagaagtccaacggcgagcggccccctgcaaccaactgagcacccgtgcagtgcaaccaaggc
 atccgccccctgggtgagcaccagctgctgctgaaacggcagcctggccgaggagggcgctgggtgatc
 cgagcgagaaacttcaccgacaacgccaagaccatcatcgtgcagctgaaggagagcgtggagatc
 aactgcacccgcccccaacaacaacaccgcaagagcatcacccatcgcccccgccgccccttctac
 gccacccggcgacatcatcgccgacatccgcccagggccccactgcaacatcagccggcgagaagtggaaac
 aacacccctgaagcagatcgtgaccaaagctgcaggcccagttcgggcaacaagaccatcgtgttcaag
 cagagcagcggcgccgagaccccgagatcgtgatgcacagcttcaactcgccggcgaggttcttctac
 tggcaacgacacccagctgtttcaacagcacctggaaacaacacccatcgcccccaacaacacacggc
 accatcacccctgcccccgccatcaagcagatcatcaaccgctggcgaggaggtgggcaaggccatg
 tacgcccccccatccgcccggccagatccgctgcagcagcaacatcacccgctcgtgctgacccgc
 gacggcggaaggagatcagcaacaccacccagagatcttccgcccccgccggcgagatcgccgagc
 aactggcgagcagcagctgtacaagtacaaggtgggtgaagatcgagccccctggcgctggcccccaac
 atcccatcagcagcgtgttgcagagcagagaagagcccgctgacccctggcgccatgttctctgggc
 tctctgggcgcccggcgagcaccatggggccccggcagcccgacccctgacccgtgcaggcccccgag
 ctgctgagcggcatcgtgcagcagcagaaacacccctgctgcccggatcgaggccccagcagcactgt
 ctgcagctgacccgtgtggggccatcaagcagctgcaggccccgctgctggccgtggagcgtctacctg
 aaggaccagcagctgctgggcatctggggctgcagcggcaagctgatctgcaccacacccgctgccc
 tggaaacgcagcctggagcaacaagagcctggaccagatctggaaacaacatgacctggatggatgg
 gagcgcgagatcgacaactacaccaaactgatctacacccctgatcgaggagagccagaaaccagcag
 gagaagaacgagcaggagctgctggagctggacaagtggccagcctgtggaaactggttcgacatc
 agcaagtggctgtggtacatccaactcgag

FIG. 33
(SEQ ID NO:46)

45/ 131

gp140.mut8.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgtgtgtggagcagtc
 ttctgtttcgccccagcgccgtggagaagctgtgggtgacctgtactacggcgtgcccgctgtggaaag
 gaggccaccaccaccctgtctctcgccagcgagcgccaaggccctacgacaccgaggtgcacaacgtg
 tggggccaccacgcctgctgccccacgaccccaacccccaggagatcgtgctggagaacgtgacc
 gagaacttcaacatgtgggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg
 gaccagagcctgaagccctgctggaagctgacccccctgtgctggggcgccggcaactgccagacc
 agcgtgataccccaggccctgccccaaaggtgagcttcgagcccatccccatccactactgcccccc
 gccggcttcgccatccctgaagtgcacacgacaagaagttcaacggcagcgcccttgccaccaacgtg
 agcacccgtgagtgacccacggcatccgccccgtggtagcaccacagctgctgtgaacggcagc
 ctggccgaggaggcgctgggtgatccgcagcgagaaacttcacggacaacgccaaagaccatcatcgtg
 cagctgaaggagagcgtggagatcaactgcacccgcccccaacaacaccccgcaagagcatcacc
 atcggcccccgccgcttctacgcccacggcgacatcatcggcgacatccgcccaggccactgc
 aacatcagccggcgaagaagtggaaacaacacccctgaagcagatcgtgaccaagctgcaggcccagttc
 ggcaacaagaccatcgtgttcaagcagagcagcgccggcgaccccgagatcgtgatgcacagcttc
 aactcgggcgggcagttctctactgcaacagcaccacagctgttcaacagcacctgggaacaacacc
 atcggcccccaacaacacacccggcacatcaccctgccctgcgcgcatcaagcagatcatcaacccg
 tggcaggagggtgggcaaggccatgtacgcccccccatccgcccagatccgctgcagcagcaac
 atcacccgctcgtgctgaccccgcgacggcggaaggagatcagcaacaccacccagatcttcgcg
 ccggcgccggcgacatgcccgaacactggcgcgagcagctgtacaagtacaagggtggtagaatc
 gagccctggcgctggccccccaccatcgccatcagcagcgtggtgcagagcggaagagcgccggtg
 accctggggcccatgttccctgggcttccctggggcgccggcgagcaccatggggcgcccgagccctg
 accctgaccctgacggcccgccagctgctgagcgccatcgtgcagcagcagaaacacctgctgcgcg
 gccatcgaggcccgagcagcacctgctgcagctgaccgtgtggggcatcaagcagctgcaggcccg
 gtgctggcgctggagcgtcactgaaggaccagcagctgctggggcatctggggctgcagcgccgaag
 ctgatctgcacacccgctgcccctggaaacgcccagctggagcaacaagagcctggaccagatctgg
 acaaacatgacctggatggagtgaggagcgcgagatcgacaactacaccaacccctgctacacccctg
 atcgaggagagccagaaccagcagggaagaacgagcaggagctgctggagctggacaagtggggc
 agcctgtggaaactggctcgacatcagcaagtggctgtggatcatctaaactcgag

FIG. 34

(SEQ ID NO:47)

gp160.modSF162

gaattcgccaccatggatgcaatgaagagagggtctgtgtgtgctgctgctgtgtggagcagtc
 ttctgtttcgccacgctgctggagaagctgtgggtgacctgtactacggcgtgcccgtgtgtggaag
 gaggccaccaccacccctgttctgtcgccagcgacgccaaggcctacgacaccgaggtgcacaaactgt
 tggggccacccacgctctgctgtcccaaccgaccccaacccccaggagatcgtgtgtggagaactgtgacc
 gagaacttcaacatgtggagaagaacaacatgggtggagcagatgcacaggacatcatcagcctgtgg
 gacagagcctgaagccctgctgaagctgacccccctgtgctgctgacccctgcactgcacaaacctg
 aagaaacgcccacacccaagagcagcaactgggaaggagatggaccgcgcgagatcaagaactgc
 agcttcaaggtgaccaccagcatccgcaacaagatgcgaaggagtagccctgttctacaagctg
 gacgtggtgcccacgcagacaacaccagctacaagctgatcaactgcacaccagcgtgatc
 acccaggcctgcccagggtgagcttcgagcccatccccatccactactgccccccgcccgttcc
 gccatcctgaagtgcacagacaagaagtccaacggcagcggccctgcaccaaactgagcaccgtg
 cagtgacccacggcatccgccccgtggtgagcaccacagctgctgctgaacggcagcctggccgag
 gaggggcgtggtgatccgcagcgagaacttcaccgcaacacgcgaagaccatcatcgtgcagctgaag
 gagagcgtggagatcaactgcacccgcccccaacaacacacccgcaagagcatcaccatcggcccc
 ggccgccccttctacgcacacggcgacatcatcggcgacatccgcccaggccccactgcaacatcagc
 ggcgagaagtggacaacacccctgaagcagatcgtgaccaaactgcaggccacgttcggcaacaag
 accatcgtgttcaagcagagcagcggcgccgaccccgagatcgtgatgcacagcttcaactgcggc
 ggcgagttcttctactgcacacgacccagctgttcaacagcacctggaaacaacacatcggcccc
 aacaacaccaaaggcacccatcaccctgcccctgcccgcacaaagcagatcatcaccgctggcaggag
 gtggggcaaggccatgtacgccccccccatccgcccagatccgctgcagcagcaacatcacccg
 ctgctgctgacccgcagcggcggaaggagatcagcaacacacccagatcttcgccccccggcgcc
 ggcgacatgcgcgacacaaactggcgacgcagctgtacaagtacaaggtggtgaagatcgagccccg
 ggctgtggcccccaacaggccaaagcgcgcgctggtgcagcgcgagaaagcgcgcctgacccctgggc
 gccatgttctctgggtctcctgggcgcgcgcggcagcaccatggggcgcgcgcagcctgacccctgac
 gtgcaggccgcagctgctgagcggcatcgtgcagcagcagaacaacactgctgcgcgcctcagag
 gccagcagcaccctgctgcagctgacccgtgtggggcatcaagcagctgcagcccccgcgtgctggcc
 gtggagcgtcactgaaggacacagcagctgctggggcatcggggctgcagcgcgaagctgatctgc
 accaccgctgcccctggaacgcgcagctggagcaacaagagcctggacagatctggaaacaacatg
 acctggatggagtgggagcgcagatcgacaactacaccaacctgatctacacctgatcgaggag
 agccagaaaccagcaggagaagaacagcagcagcagctgctggagctggacaagtggggcagcctgtgg
 aactggttcgacatcagcaagtggctgtggtacatcaagatcttcatcatgatcgtgggcgcgcctg
 gtgggctgcgcactgtgttccagctgctgagcctggaacccgctgcccgcgcgcagggctacagcccc
 ctgagcttccagacccccttccccgcccccgccgccccgaccccccaggggcatcgaggaggag
 ggccggcagcgcagccgcagccgcagcagcccccgtgtgcacggcctgctggccctgatctggga
 gacctgcgcagcctgtgcccgttccagctacccacgcctgcgcgacctgatcctgatcgccgcgcgc
 atcgtggagctgctgggcgcgcgcgcctgggaggccctgaagtactggggcacaactgctgcagatc
 tggatccaggagctgaagaacagcgcgcctgagcctgttcagcgcacatcgccatcgccgtggccgag
 ggcaaccgacgcacatcagaggtggccacagcgcacggccgcgccttctgcacatccccgcgcg
 atccgcagaggcttcagagcgcgcctgctgttaactcgag

FIG. 35
(SEQ ID NO:48)

gp160.modSF162.delV2

gaattcgccaccatggatgcaatgaagagaggggctctgctgtgtgctgctgctgtgtggagcagtc
 ttcgttttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgctgccctgtggaag
 gagggccaccaccacctgttctgcgccagcgacgccaaggcctacgacacccagaggtgcacaacgtg
 tggggccaccacgctcgctgccaccgaccccaacccccaggagatcgtgctggagaacgtgacc
 gagaacttcaacatgtggagaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg
 gaccagagcctgaagccctcgctgaaagctgacccccctgtgctgacccctgcactgcaccaacctg
 aagaacgcccaacaccaaagcagcgaactggaaggagatggaccgcggcgagatcaagaacctgc
 agcttcaagggtggggcgccggcaagctgatcaactgcaacaccagcgtgatcacccaggcctgcccc
 aaggtgagcttcgagcccatccccatccactactgcgccccccgcccgttcgcccactcctgaagtgc
 aacgcacaagaagtccaacggcgagcggccctgcaccaacgtgagcacccgtgcagtgcaccacacggc
 atccgccccctggtagcacccagctgctgctggaacggcagcctggccgaggaagggcggtggtgatc
 cgacgagagaacttcaccgacaaacgcgaagacatcatcgtgcagctgaaggagagcgtggagatc
 aactgcacccgcccccaacaacaccccgcaagagcatcaccatcgcccccggcgccgcttctac
 gccaccggcgacatcatcgccgacatccgccaggcccaactgcaacatcagcggcgagaagtggaaac
 aacacccctgaagcagatcgtgaccaagctgcaggcccaactgcggcaacaagaccatcgtgttcaag
 cagagcagcggcgccgaccccgagatcgtgatgcacagcttcaactgcggcgccgagattctctac
 tgcacaagcagccccagctgttcaacagcacctggaacaacacccatcgcccccacaacacacacggc
 accatcacccctgccctgcgcgatcaagcagatcatcaacgcctggcaggaggtgggcaagggccatg
 tacgccccccccctccggccagatccgctgcagcagcaacatcacccgctgctgctgacccgc
 gacggcggaagagatcagcaacacccacagatcttcgcccccgccggcgccagatcgcgacac
 aactggcgagcgagcgtgtacaaagtacaaggtgggtgaagatcgagccccctggcgctggccccacac
 aaggccaagcgcccgctggtgcagcgaggaagcgcccgctgacccctggcgccatgttccctgggc
 ttcctggggcgcccgccgagcaccatggggcgcccgagcctgacccctgacccgtgcaggcccgccag
 ctgctgagcggcatcgtgcagcagcagaacaacccctgctgcgcgccatcgaggccccagcagacctg
 ctgcagctgacccgtgtggggcatcaagcagctgcaggcccgctgctggccgtggagcgtacctg
 aaggaccagcagctgctgggcatctggggctgcagcggcaagctgatctgcacacccgccgtgcc
 tggaaacgccagctggagcaacaagagcctggaccagatctggaaacaacatgacctggatggagtgg
 gagcgcgagatcgacaactacaccaacctgatctacacccctgatcgaggagagccagaaccagcag
 gagaagaacgagcaggagctgctggagctggacaagtggggcagcctgctggactggttcgacatc
 agcaagtggtgtggtacatcaagatcttcatcatgatcgtggggcgccctgggtgggctgcgcac
 gtgttcacgctgctgagatcgtggaacgcgctgcggcagggtacagccccctgagcttccagac
 cgtgtcccccggcccgcccgacccgaccccgaggggcatcgaggaggaggggcgcgagcgcgac
 cgcgaccgcagcagccccctgggtgcacggcctgctggccctgatctgggaacgacctgcgcagcctg
 tgcctgttcagctaccacccgctgcgcgacatgatcctgatcgccggccgacatcgtggagctgctg
 ggccgcccggctggggggccctgaagtactggggcaacctgctgcagtaactggatccaggagctg
 aagaacagcgcctgagcctgttcgacgcctatcgccatcgccgtggcgaggggcaacgacccgac
 atcgaggtggccagcgcatcgccgcgccttctgcacatcccccgccgatccgcccagggtctc
 gagcgcgcccctgctgtactcgag

FIG. 36
 (SEQ ID NO:49)

gp160.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc
 ttcgtttcgccagcgcctggagaagctgtgggtgacgctgtactacggcgtgcccgttggaa
 gaggccaccaccacccctgttctcgccgacgacgccaagggctacgacaccgaggtgcacaacgtg
 tggggcaccaccagcctgcgtgcccaccgaccccaaccccgagagatcgtgctggagaacgtgacc
 gagaacttcaacatgtgggaagaacaacatggaggagcagatgcacgaggaatcatcagcctgtg
 gaccagagcctgaagccctgcgtgaagctgacccccctgtgctggggcgcggcaactgccagacc
 agcgtgatcacccaggcctgcccacaggtagcttcgagcccatccccatccactactgcgcgcgc
 gccggcttcgccatcctgaagtgcacacgacaagaagtccaacggcagcggcccttgaccaacgtg
 agcaccgtgcagtgacccacggcatccgcccgtggtagaccccgagctgctgctgaacggcagc
 ctggccgaggaggcgtggtgatccgcagcgaagaacttcaccgcacaacgcgaagaccatcatcgtg
 cagctgaaggagagcgtggagatcaactgcacccgccccacaacaacacccgcgaagagcatcac
 atcggccccggcgcgccttctacgcccacggcgacatcatcggcgacatccgcccaggccactgc
 aacatcagcggcgagaagtgggaacaacacccctgaagcagatcgtgaccaagctgcaggccaggtt
 ggcaacaagaccatcgtgttcaagcagagcagcggcgaggcagaccccgagatcgtgatgcacagctt
 aactgcggcgcgagttcttactgcacacgaccccgagctgttcaacagcacctgggaacaacacc
 atcggccccacaacaacacacggcaccatcacccctgccctgccgcatcaagcagatcatcaacgc
 tggcaggagggtgggcaaggccatgtacgccccccccatccgcggccagatccgctgcagcagcaac
 atcacccggcctgctgctgacccgcgacggcggaaggagatcagcaaacacacccagagatcttcgc
 cccggcgcgcgagcatgcccgaacactggcgacgagcgtgtacaagtaacaagggtggtgaagatc
 gagccctgggctggcccccccaaggccaaagcgcgcgcgtggtgcagcgcggaagcgcgcgcgtg
 accctgggcgcctggttccctgggcttccctgggcgcccgccgagcaccatgggcgcgcgcgcgcgtg
 accctgacctgcgcggcccgccagctgctgagcggcatcgtgcagcagcaacaacactgctgcgc
 gccatcgaggcccgagcagccctgctgacagctgacagctgtggggcatcaagcagctgcaggccgcg
 gtgctggcgtggagcgcctacctgaaggaccagcagctgctgggcatctggggctgcagcgcgaag
 ctgatctgcaccaccccgctgcccgggaacgcgcagctggagcaacaagagcctggaccagatctgg
 aacaacatgacctggatggagtgggagcgcgagatcgacaactacaccaacactgacatcacccctg
 atcgaggagagccgaacaacagcaggagaagaacgagcaggagctgctggagctggacaagtgggc
 agcctgtggaaactggttcgacatcagcaagtggtgctggtacatcaagatcttcatcatgatcgtg
 ggcgccctggtggcctgcgcacatcgtgttcacgctgctgagcatcgtgaacccgcgtgcgcgcagg
 taccgccccctgagcttcacgaacccgcttccccgcgccccgcggccccgacccgccccgagggc
 gaggaggaggggcggcgagc
 atctgggacgacctgcgcagcctgtgcctgttcagctaccacccgctgcgcgcgcgcgcgcgcgcgc
 gccgcccgcacatcgtggagctgctgggc
 ctgcagctactggatccaggagctgaagaacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
 gtggcgagggcaccgaacgcacatcgagggtggccagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
 cccgcgcgcacccgcagggttcgagc

FIG. 37
 (SEQ ID NO:50)

gp120wtUS4

ACAACAGTCTTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAAGAAG
CAACCACCACTCTGTTTTGTGCATCAGATGCTAAAGCATACAAAGCAGAGGC
ACATAACGTCTGGGCTACACATGCCTGTGTACCCACAGACCCCAACCCACAG
GAAGTAAATTTAACAAATGTGACAGAAAATTTTAAACATGTGGAAAAATAACA
TGGTGGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAAGCCTAAA
GCCATGTGTAAAATTAACCCCACTCTGTGTTACTTTAAATTGTACTGATAAGT
TGACAGGTAGTACTAATGGCACAATAAGTACTAGTGGCACTAATAGTACTAG
TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGGAAAAGATGCCA
GAAGGAGAAATAAAAACTGCTCTTTCAATATCACCACAAGTGAAGAGATA
AAGTGCAGAAAATAATTCTCTCTCTATAAACTTGATGTAGTACCAATAGAT
AATGATAATGTAGCTATAGATTGATAAATTGTAATACCTCAGTACATTACACA
AGCCTGTCCAAAGGTATCTTTTGAACCAATTTCCCATACATTATTGTGCCCGG
CTGGTTTTGCGATTCTAAAGTGTAAGATAAGAAGTTCAATGGAAACAGGACC
ATGTAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAGTAGTA
TCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA
GATCTGAAAAATTTACAGACAATGTAAAAACCATAATAGTACAGCTGAATGA
ATCTGTAGAAAATTAATTGTATAAGACCCCAACAATAACAAGAAAAAGTATA
CATATAGGACCAGGGAGAGCATTTTTATGCAACAGGTGATATAATAGGAGACA
TAAGACAAGCACATTGTAACATTAGTAAAGCAAACTGGACTAACACTTTAGA
ACAGATAGTTGAAAAATTAAGAGAACAATTTGGGAATAATAAAACAATAATC
TTTAATTCATCCTCAGGAGGGGACCCAGAAAATTGTATTTACAGTTTAAATTG
TGGAGGGGAATTTTCTATTGTAATACATCACAACTATTTAATAGTACCTGGA
ATATTACTGAAGAGGTAATAAGACTAAAGAAAATGACACTATCATACTCCC
ATGCAGAATAAGACAAAATTATAAACATGTGGCAAGAAGTAGGAAAAGCAAT
GTATGCCCTCCCATCAGAGGACAAATTAATGTTCATCAAAATATTACAGGG
CTGCTATTAAGTAGAGATGGTGGTACTAACAATAATAGGACGAACGACACCG
AGACCTTCAGACCTGGGGGAGGAAACATGAAGGACAATTGGAGAAGTGAAT
TATATAATATAAAGTAGTAAGAAATTGAACCATTAGGAGTAGCACCCACCCA
GGCAAGAGAGAAGAGTGGTGCAAGAGAGAGAAAAGA

FIG. 38
(SEQ ID NO:51)

50 / 131

gp140wtUS4

ACAACAGTCTTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAAGAAG
CAACCACTCTGTTTTGTGCATCAGATGCTAAAGCATACAAGCAGAGGC
ACATAACGTCTGGGCTACACATGCCTGTGTACCCACAGACCCCAACCCACAG
GAAGTAAATTTAACAAATGTGACAGAAAAATTTAACATGTGGAAAAATAACA
TGGTGGAAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAA
GCCATGTGTAAAATTAACCCCACTCTGTGTTACTTTAAATTTGACTGATAAGT
TGACAGGTAGTACTAATGGCACAATAGTACTAGTGGCACTAATAGTACTAG
TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGGAAAAGATGCCA
GAAGGAGAAATAAAAACTGCTCTTTCAATATCACCACAAGTGTAAAGAGATA
AAGTGCAGAAAGAATATTCTCTCTCTATAAACTTGATGTAGTACCAATAGAT
AATGATAATGCTAGCTATAGATTGATAAATTTGTAATACCTCAGTCATTACACA
AGCCTGTCCAAGGATCTTTTGAACCAATTCCCATACATTATTGTGCCCCGG
CTGGTTTTGCGATTCTAAAGTGTAAAGATAAGAAGTTCAATGGAAACAGGACC
ATGTAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAGTAGTA
TCAACTCAACTGCTGTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA
GATCTGAAAAATTCACAGACAATGCTAAAAACCATAATAGTACAGCTGAATGA
ATCTGTAGAAATTAATTGTATAAGACCCAAACAATAATACAAGAAAAAGTATA
CATATAGGACCAGGGAGAGCATTTTATGCAACAGGTGATATAATAGGAGACA
TAAGACAAGCACATTGTAAACATTAGTAAAGCAAACTGGACTAAACCTTTAGA
ACAGATAGTTGAAAAATTAAGAGAACAATTTGGGAATAATAAAACAATAAT
TTTAATTCATCCTCAGGAGGGGACCCAGAAATTTGATTTTACAGTTTTAATTG
TGGAGGGGAATTTTTCTATTGTAATACATCACAACTATTTAATAGTACCTGGA
ATATTACTGAAGAGGTAAATAAGACTAAAGAAAAATGACACTATCATACTCCC
ATGCAGAATAAGACAATTTATAAACATGTGGCAAGAAGTAGGAAAAAGCAAT
GTATGCCCTCCCATCAGAGGACAAATTAATGTTTCATCAAAATATTACAGGG
CTGCTATTAAGTACAGATGGTGGTACTAACAATAATAGGACGAACGACACCG
AGACCTTCAGACCTGGGGGAGGAAACATGAAGGACAATTTGGAGAAGTGAAT
TATATAAATATAAAGTAGTAAGAATTGAACCATTAGGAGTAGCACCACCA
GGCAAAAGAGAAAGAGTGGTGCAAAAGAGAGAAAAAGCAGTGGGACTAGGAG
CTTTGTTCATTGGGTTCTTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTC
AGTGACGCTGACGGTACAGGCAGACAATTAATGTCTGGTATAGTGAACAG
CAGAACAAATTTGCTGAGAGCTATTGAGGCGCAACAGCATCTGTGCAACTCA
CGGTCTGGGGCATCAAACAGCTCCAGGCAAGAATCCTGGCTGTGAAAAAGATA
CCTAAAGGATCAACAGCTCCTAGGGAATTTGGGGTTGCTCTGGAAAACTATT
GCACCACTACTGTGCTTGGAACTCTAGTTGGAGTAATAAATCTCTGACTGAG
ATTTGGGATAAATGACCTGGATGGAGTGGGAAAGAGAAATTTGGCAATTATA
CAGGCTTAATATACAATTTAATTGAAATAGCAAAAAACCAGAAAGAAAAAGA
TGAACAAGAATTATTGGAATTAGACAAGTGGGCAAGTTTGTGGAATTGTTT
GATATAACAACTGGCTGTGGTATATA

FIG. 39

(SEQ ID NO:52)

51 / 131

gp160wtUS4

ACAACAGTCTTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGAAAGAAG
CAACCACCACTCTGTTTTGTGCATCAGATGCTAAAGCATACAAAGCAGAGGC
ACATAACGTCTGGGTCACACATGCCTGTGTACCCACAGACCCCAACCCACAG
GAAGTAAATTTAACAAATGTGACAGAAAATTTTAAACATGTGGAAAAATAACA
TGGTGGAAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAA
GCCATGTGTAAAAATTAACCCCACTCTGTGTTACTTTAAATGTGACTGATAAGT
TGACAGGTAGTACTAATGGCACAATAGTACTAGTGGCACTAATAGTACTAG
TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGGAAAAGATGCCA
GAAGGAGAAAATAAAAACTGCTCTTTCAATATCACCACAAGTGTAAAGAGATA
AAGTGCAGAAAAGAATATTCTCTCTTCTATAAACTTGATGTAGTACCAATAGAT
AATGATAATGCTAGCTATAGATTGATAAATTGTAATACCTCAGTCATTACACA
AGCCTGTCCAAAGGTATCTTTTGAACCAATTCCCATACATTATTGTGCCCCGG
CTGGTTTTGCGATTCTAAAGTGTAAGATAAGAAGTTCAATGGAACAGGACC
ATGTAAAAATGTGCAGCAGTACAATGCACACATGGAATTAGACCAGTAGTA
TCAACTCAACTGCTGTGTTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA
GATCTGAAAAATTTACAGACAATGCTAAAACCAATAATAGTACAGCTGAATGA
ATCTGTAGAAATTAATTGTATAAGACCCAAACAATAACAAGAAAAAGTATA
CATATAGGACCGGAGAGCATTTTATGCAACAGGTGATATAATAGGAGACA
TAAGACAAGCACATTGTAAACATTAGTAAAGCAAACCTGGACTAACCTTTAGA
ACAGATAGTTGAAAAATTAAGAGAACAAATTTGGGAATAATAAAACAATAATC
TTTAATTCCTCTCAGGAGGGGACCCAGAAATTTGATTTTACAGTTTTAATTG
TGGAGGGGAATTTTTCTATTGTAATACATCACACTATTTAATAGTACCTGGA
ATATTACTGAAGAGGTAAATAAGACTAAAGAAAATGACACTATCATACTCCC
ATGCGAATAAAGACAATTATAAAACATGTGGCAAGAAGTAGGAAAAAGCAAT
GTATGCCCTCCCATCAGAGGACAAATTAATGTTTCATCAATATTACAGGG
CTGCTATTAACTAGAGATGGTGGTACTAACAATAATAGGACGAACGACAGG
AGACCTTCAGACCTGGGGGAGGAAACATGAAGGCAATTTGAGAAAGTGAAT
TATATAAATATAAAGTAGTAAGAATTGAACCATTAGGAGTAGCACCCACCCA
GGCAAGAGAGAAGAGTGGTGCAAAGAGAGAAAAAGAGCACTGGGACTAGGAG
CTTGTGTTCTATGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTC
AGTGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAACAG
CAGAACAAATTTGCTGAGAGCTATTGAGGCGCAACAGCATCTGTTGCAACTCA
CGGCTCTGGGGCATCAAAACAGCTCCAGGCAAGAATCCTGGTCTGGAAGATA
CCTAAAGGATCAACAGCTCCTAGGGATTGGGGTGTCTGGAATACTCATT
GCACCACTACTGTGCTTGGAACTCTAGTTGGAGTAATAAATCTCTGACTGAG
ATTTGGGATAATATGACCTGGATGGAGTGGGAAAGAGAAAATTGGCAATTATA
CAGGCTTAATATACAATTTAATTGAAATAGCAAAAACCGCAAGAAAAGAA
TGAACAAGAAATTATTGGAATTAGACAAGTGGGCAAGTTTGTGGAATTGGTTT
GATATAACAAACTGGCTGTGGTATATAAGAATATTCAATGATAGTAGGAG
GCTTGATAGGTTTAAAGAATAGTTTTTGTCTGACTTTCTATAGTGAATAGAGTT
AGGCAGGGATACTACCAATATCATTGCAGACCCGCCCTCCAGCTCAGAGGG

FIG. 40A

(SEQ ID NO:53)

52 / 131

GACCCGACAGGCCCGAAGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGA
GACAGATCCAATCGATTAGTGCATGGATTATTGGCACTCATCTGGGACGATCT
GCGGAGCCTGTGCCTCTTCAGCTACCAACCGCTTGAGAGACTTACTCTTGATTG
TAGCGAGGATTGTGGAATTCTGGGACGCAGGGGGTGGGAAGCCCTCAAGTA
TTGGTGGAATCTCCTGCAGTATTGGAGTCAGGAGCTAAAGAGTAGTGCTGTT
AGTTTGTTTAATGCCACAGCAATAGCAGTAGCTGAAGGGACAGATAGGATTA
TAGAAATAGTACAAAGAATTTTATAGAGCTGAATTCACATACCTAGAAGAAT
AAGACAGGGCTTGGAGAGGGCTTTACTATAA

FIG. 40B

(SEQ ID NO:53)

gp120.modUS4

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGTGTGGAGCA
 GTCTTCGTTTCGCCACGCCACCGTGTGTGGGTGACCGTGTACTACGGCGTGCCCGTG
 TGGAAAGGAGGCCACCAACCCTGTTCTGGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC
 CCACAAACGTGTGGGGCCACCCACGCCCTGCGTGCCACCCGACCCCAACCCCCAGGAGGTGAAC
 TGACCAACGTGACCGAGAACTTCAACATGTGGAAGAAACAATGGTGGAGCAGATGCATGAG
 GACATCATCAGCCTGTGGGACCAGAGCCTGAAAGCCTGCGTGAAGCTGACCCCCCTGTGCGTG
 ACCCTGAACTGCACCGACAAGCTGACCGGACGCCAACCGGCCAACAGCACCGCGGCAC
 CAACAGCACCGCGGCCAACAGCACCGACGCCAACAGCACCGACAGCTGGGAGAAAGATG
 CCCGAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCAACAGCGTGCAGCAAGGTGCA
 GAAGGAGTACAGCCTGTTCTCAAGCTGGACGTGGTGCCCATCGACAAAGCAACCGCCAGCT
 ACCGCTGATCACTGCAACACCGCGTGATCAACCGGCTGCCCAAGGTGAGCTTCGAGC
 CCATCCCCATCCACTACTGCGCCCCCGCGGCTTGCCATCCTGAAGTGCAAGGACAAGAAAT
 TCAACGGCAACCGGCCCTGCAAGAAAGTGAAGACCGTGACGTGCAACCGCATCCGCCCC
 GTGGTGAGCACCCAGCTGTGTGTAACGGCAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTC
 CGAGAACTTACCGACAACGCCAAGACCATCATCGTGACGTGAACAGAGTCCGTGGAGATCA
 ACTGCATCCGCCCAACAACAACACCGGTAAAGAGCATCCACATCGGCCCGCGCCGCGCTTCT
 ACGCCAACGGCGACATATCGGCGACATCGGCCAGGCCACTGCAACATCAGCAAGGCCAAC
 TGGACCAACCCCTCGAGCAGATCGTGGAGAAAGTGCAGCAGCATGTCGGAACAAAGAC
 CATCATCTTCAACAGCAGCAGCGCGCGCGACCCGAGATCGTGTTCCACAGCTTCACTGCGG
 CGCGAGTTCTTCTACTGCAACACCGACCGAGCTGTTCACAGCACTGGAACATCACCGAGGA
 GTGTAACAAGACCAAGGAGAAAGCAACCATCATCCTGCGCTGCCGATCCGCCAGATCATCA
 ACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCGCCCATCCGCGGCCAGATCAAGTGC
 AGCAGCAATATACCGGCTGTGTGTGACCCGCGCAGCGCGGCAACCAACAACCGCACCA
 CGACCGGAGACCTTCGCGCCCGCGCGCGCAACATGAAGGACAACTGGCGCAGCGAGCTGT
 ACAAGTACAAGGTGTGCGCATCGAGCCCTGGGCGTGGCCCCACCCAGGCCAAGCGCCG
 GTGGTGACGCGGAGAAAGCGCTAAGATATCGGATCCTCTAGA

FIG. 41
(SEQ ID NO:54)

gp120.mod.US4.del128-194

GAATTCCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGG
AGCAGTCTTCGTTTCGCCCAGCGCCACCACCGTGTGTGGGTGACCGTGTACTACGGCG
TGCCCGTGTGGAAGGAGGCCACCACCCTGTTCTGCGCCAGCGACGCCAAGGCTTAC
AAGGCCGAGGCCCAACAGTGTGGGCCACCCACGCTGCGTGCCCACCGACCCCAACCC
CCAGGAGGTGAACCTGACCAACGTGACCCAGAACTTCAACATGTGGAAGAACAACATGG
TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG
AAGCTGACCCCCCTGTGCGTGGGGCAGGGAAGTGCAGAGACCAGCGTGATCACCAGGC
CTGCCCCAAGGTGAGCTTCGAGCCCATCCCATCCACTACTGCGCCCCCGCGGCTTCG
CCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGCCCTGCAAGAAGCTGAGC
ACCGTGCAGTGCACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGG
CAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACTTCACCGACAACGCCAAGA
CCATCATCGTGCACTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCAACAACAAC
ACGCGTAAGAGCATCCACATCGGCCCGGCGCGCCTTCTACGCCACCGCGCATCAT
CGCGCATATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAAGTGGACCAACCCCTCG
AGCAGATCGTGGAAGCTGCGCGAGCAGTTCGGCAACAACAAGACCATCATCTTCAAC
AGCAGCAGCGCGCGGACCCCGAGATCGTGTTCACAGCTTCAACTGCGCGCGCGAGTT
CTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAAACATCACCGAGGAGGTGA
ACAAGACCAAGGAGAACGACACCATCATCTGCCCCTGCGGCATCCGCCAGATCATCAAC
ATGTGGCAGGAGTGGCAAGGCCATGTACGCCCCCCCATCCGCGGCCAGATCAAGTG
CAGCAGCAATATTACGGCCTGCTGCTGACCCGCGAGCGGCAACATGAAGGACAACCTGGCGCAGC
CCAACGACACCGAGACCTTCGCCCCGGCGGCGCAACATGAAGGACAACCTGGCGCAGC
GAGCTGTACAAGTACAAGTGGTGCGCATCGAGCCCTGGGCGTGCCCCCAACCCAGGC
CAAGCGCCGCGTGGTGACGCGGAGAAGCGCTAAGATATCGGATCCTCTAGA

FIG. 42

(SEQ ID NO:55)

55 / 131

gp140.modUS4

GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA
GTCCTCGTTTCGCCACGCGCCACCAACCGTGCTGTGGGTGACCGTGTACTACGGCGTGCCCGTG
TGGAAGGAGGGCAACCAACCGTCTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGG
CCACAACGTGTGGGCCACCCACGCTGCGTGCCACCGACCCCAACCCCGAGGAGGTGAACC
TGACCAACCGTGACCGAGAACTTCAACATGTGGAAAGAACATGGTGGAGCAGATGCATGAG
GACATCATCAGCCTGTGGGACCGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG
ACCTGAACTGCAACGACAAAGTGAACCGGACGACCAACCGGACCAACGACACCGGAC
CAAACGACACCGCGGACCAACAGCAACGACCAACAGCAGCAGCAGCTGGGAGAAAGATG
CCCGAGGGCGAGATCAAGAACTGCAAGCTTCAACATCAACACCGAGCTGCGCGACAAGGTGCA
GAAGGAGTACAGCCTGTTCTAAGCTGGAGCTGGTGCCATCGACAACGACAACCGCAGCT
ACCGCTGATCAACTGCAACACCGCTGATCAACCGGCTGCCCCAAGGTGAGCTTCGAGC
CCATCCCCATCCACTACTGCGCCCCGCGGCTTCGCCATCCTGAAAGTCAAAGGACAAGAAAT
TCAAACGGCACCGGCCCTGCAAGAACTGAGCAGCTGCAAGTGAACCCACGCGATCCGCCCC
GTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCTGGCCGAGGAGGAGATCGTGCTGCGCTC
CGAGAACTTCAACGACAACGCAAGCAACCATCATCTGTGACGTGAACGAGTCCGTGGAGATCA
ACTGCAATCCGCCCAACAACAACACCGCTAAGAGCATCCACATCGGCCCGCGCGCTTCT
ACGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAAC
TGGACAACAACCTCGAGCAGATCGTGGAGAAAGCTGCGCGAGCAGATTGCGCAACAACAAGAC
CATCATCTTCAACAGCAGCAGCGCGGCGGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGG
CGCGAGTTCCTTCTACTGCAACACCGACCGAGCTGTTCAACAGCACCTGGAACATCAACCGAGGA
GGTGAAACAAGCAAGGAGAAACGCAACCATCATCTGCCCTGCGCGATCCGCCAGATCATCA
ACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCGCCCATCGCGGCCAGATCAAGTGC
AGCAGCAATATTACCGGCTGCTGCTGACCCGCGACGGCGGCAACCAACAACCGCAACCA
CGACACCGAGACCTTCGCCCGCGCGGCGGCAACATGAAGGACAACCTGGCGAGCGAGCTGT
ACAAGTACAAGGTGGTGCAATCGAGCCCTGGGCGTGGCCCCACCGAGGCAAGCGCGCG
GTGGTGAGCGCGAGAAAGCGCGCTGGGCGCTGGCGCGCTGTTTATCGGCTTCTGGGCGCG
GCCGGGAGCACCATGGGCGCGCTCCGTGACCCGTGACCGGTGACGGCCCGCGCAGCTGCTGAG
CGGCATCGTGACGACGAGAACCAACCTGCTGCGCGCAATCGAGGCCACGACGACCTGCTGC
AGCTGAACCGTGTGGGCGATCAAGCAGCTGACGGCCCATCTGGCCGTGGAGCGCTACCTG
AAGGACGAGCAGCTGTGGGATCTGGGCGTGCAGCGGCAAGCTGATCTGACCAACCAACCTG
GCCTGGAAACAGCAGCTGGAGCAACAAGAGCTGACCGAGATCTGGGACAACATGACCTGGA
TGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCTGATCTACAACCTGATCGAGATCGCC
CGAACAACAGAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCGACGCTGT
GGAAGTGTTCGACATCAACAACCTGGCTGTGGTACATCTAAGATATCGGATCCTCTAGA

FIG. 43

(SEQ ID NO:56)

gp140.mut.modUS4

GAATTGCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTCTGCTGCTGTGTGGAGCA
 GTCTTCGTTTCGCCCACGCCACCCACCGTGTGTGGGTGACCGTGTAACCGGGCTGCCGTG
 TGGAAAGGAGGCCACCAACCCCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC
 CCAACAACGTGTGGGCCACCCACGCTGCGTGCCACCGACCCCAACCCAGGAGGTGAACC
 TGACCAACGTGACCGGAGAACTTCAACATGTGGAAAGAACAACATGGTGAGCAGATGCATGAG
 GACATCATCAGCCTGTGGGACCAAGCCTGAAGCCCTGCGTGAAGCTGACCCCTGTGCGTG
 ACCCTGAACGTGACCGACAAGCTGACCGGACGACCAACGGCACCAACAGCACACGGGCAC
 CAACAGCAACGAGCGGACCAACAGCACCAAGCAACAGCACCGGACAGCTGGGAGAAGATG
 CCCGAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCAAGCGTGCAGCAACAGGTGCA
 GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCATCGACAACGACAACGCCAGCT
 ACCGCTGTATCAACTGCAACACCAAGCGTGATCAACCAAGCCTGCCCAAGGTGAGCTTCGAGC
 CCAATCCCATCCAATACTGCGCCCCCGCCGGCTTCGCCATCTGAAAGTGAAGGACAAGAAGT
 TCAACGGCAACGGCCCCCTGCAAGAACGTGAGCACCGTGCAAGTGCACCCACGGCATCCGCCCC
 GTGGTGAGCAACCGAGCTGTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCTGTGCTGCGCTC
 CGAGAACTTCAACGACAACGCCAAGACCATCATCTGTGACGTGAACGAGTCCGTGGAGATCA
 ACTGCATCCGCCCAACAACAACAGCGTAAGAGCATCCACATCGGCCCGCGCCGCGCTTCT
 ACGCCACCGGCACATCATCGCGCATCCGCCAGGCCCACTGCAACATCAAGCAAGGCCAAC
 TGGACCAACACCCCTCGAGCAGATCGTGAGAAAGCTGCGGAGCAAGTTCGGAACAACAAGAC
 CATCATCTTCAACAGCAGCAGCGCGCGGCAACCCGAGATCGTGTCCACAGCTTCAACTGCGG
 CGCGGAGTTCTTCTACTGCAACACCAAGCCAGCTGTTCAACAGCACCTGGAACATCAACCGAGGA
 GGTGAACAAGACCAAGGAGAAACGACCATCATCTGCTGCCCTGCGGCAACCGGCAAGATCA
 ACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCATCCCGGCCAGATCAAGTGC
 AGCAGCAATATTACCGGCTGTGCTGACCGCGACGGCGGCAACCAACAACACCGCACCA
 CGACACCGAGACCTTCCGCCCCGCGCGGCAACATGAAGGACAACCTGGCGCAGCGAGCTGT
 ACAAGTACAAGGTGGTGCATCGAGCCCTGGGCGTGGGCCCTGTTTCACTCGGCTTCCTGGGCGC
 GTGGTGACGCGGAGAAAGAGCGCCGTGGGCTGGGCGCCCTGTTTCACTCGGCTTCCTGGGCGC
 GCCGGGACACCATGGGCGCCGCTCCGTGACCTGACCGTGACGGCCCGCCAGCTGCTGAG
 CGGCATCTGTGCAGCAGCAACAACCTGCTGCGCGCATCGAGGCCAGCAGCACCTGCTGC
 AGCTGACCGTGTGGGCAATCAAGCAGCTGACAGGCCCGCATCTCTGGCGTGGAGCGCTACCTG
 AAGGACCAAGAGCTGCTGGGCACTCTGGGCTGACAGCGGCAAGCTGATCTGCAACCAACCGT
 GCCCTGGAACAAGCAGCTGGAGCAACAAGAGCCTGACCGGAGATCTGGGACAACATGACCTGGA
 TGGAGTGGGAGCGGAGATCGGCAACTACACCGGCTGATCTACAACCTGATCTGAGATCGCC
 CAGAACCAAGAGGAGAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGT
 GGAACCTGTTGACATCAACCACTGGCTGGGTACATCTAAGATATCGGATCTCTAGA

FIG. 44

(SEQ ID NO:57)

57 / 131

gp140.TM.modUS4

GAATTGCGCCACCATGGATGCAATGAAAGAGAGGGGCTCTGCTGTGTGCTGCTGTGTGGAGCA
GTCTTCGTTTCGCCACGCGCCACCAACCGTGCTGTGGGTGACCGTGTACTACGGCGTGCCCGTG
TGGAAAGGAGGCCACCAACACCTGTTCTGCGCCACGCAACGCCAAGGCTTACAAGGCCGAGGC
CCACAAACGTGTGGGCCACCCACGCTGCGTGCCACCGACCCCAACCCCGAGGAGGTGAACC
TGACCAACGTGACCGGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCAATGAG
GACATCATCAGCCTGTGGGACCAAGCTGACCGGACGACCAACGCGACCAACAGCACCAGCGGCAC
ACCTTGAACCTGCACCGACAAGCTGACCGGACGACCAACGCGACCAACAGCACCAGCGGCAC
CAAACAGCACCAGCGGCACCAACAGCACCAGCACAACAGCAGCAGCAGCAGCAGCAGCAGCAGC
CCGAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCACAGCGTGGCGGACAAAGGTGCA
GAAGGAGTACAGCCTGTTCTCAAGCTGGAGCTGGTGCCCATCGACAACGACAACGCCAGCT
ACCGCTGATCAACTGCAACACCAGCGTGATCACCCAGGCTGCCCAAGGTGAGCTTCGAGC
CCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAGT
TCAAACGCGACCGGCCCTGCAAGAACGTGAGCACCGTGCAATGCAACCAACGCGCATCCGCCCC
GTGGTGAGCAACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTC
CGAGAACTTCAACGACAACGCCAAGACCATCATCGTGACGCTGAACGAGTCCGTGGAGATCA
ACTGCATCCGCCCAACAACAACACGCGTAAGAGCATCCACATCGGCCCGCCCGCCGCTTCT
ACGCCACCGCGCATCATCGCGCATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC
TGGACCAACACCTCGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCGCGAACAACAAGAC
CATCATCTTCAACAGCAGCAGCGCGCGGACCCCGAGATCGTGTTCACAGCTTCAACTGCGG
CGCGGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACTGGAAACATCAACGAGGA
GGTGAACAAGCAAGGAGAACGACACCATCATCTGCGCTGCGCATCCGCCAGATCATCA
ACATGTGGCAGGAGGTGGGCAAGCCATGTACGCCCCCCTCGCGCGGCGAGATCAAGTGC
AGCAGCAATATTACCGGCTGCTGCTGACCCGCGCAGCGCGGACCAACAACACCGCAACCA
CGACACCGAGACCTTCGCCCCGCGCGGCGCAACATGAAGGACAACCTGGCGAGCGAGCTGT
ACAAGTACAAGGTGGTGCGCATCGAGCCCCCTGGGCGTGGCCCCACCCAGGCCAAGCGCGCG
GTGGTGACGCGGAGAAAGCGCGCCGTGGGCGCTGGGCGCCCTGTTTCATCGGCTTCTGGGCGGC
CGCGGAGCAACCATGGGCGCGCTCCGTGACCCCTGACCGTGACGCGCGCGCAGCTGCTGAG
CGGCATCGTGACGAGCAGAGAAACCTGCTGCGCGCATCGAGGCCAGCAGCACTGCTGC
AGCTGACCGGTGTGGGCACTCAAGCAGCTGCAACGCGCAAGCTGATCTGCACCAACCAACCGT
AAGGACAGCAAGCTGCTGGGCTGCAACGAGCTGACCGAGATCTGGGACAACATGACCGCTACCTG
GCCCTGGAACAGCAGCTGGAGCAACAAGACGCTGATCTGCAACCTGATCGAGATCGCC
TGGAGTGGGAGCGGAGATCGGCAACTACACCGGCTGATCTCAACCTGATCGAGATCGCC
CAGAACCGAGCAAGGAGAAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCTGT
GGAACGTGTTGACATCAACCACTGGCTGTGGTACATCCGATCTTCAATCATGATCGTGGGCG
GCCTGATCGGCTGCGCATCGTGTTCGCGCTGCTGAGCATCGTGTGAAGATATCGGATCTCTA
GA

FIG. 45
(SEQ ID NO:58)

Gp140modUS4.DV1V2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGC
TGTGTGGAGCAGTCTTCGTTTCGCCACGCGCCACCAACCGTGTGTGGGTGACC
GTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCAACCCCTGTTCGCG
CCAGCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTGGGCCACCCA
CGCTCGCTGCCACCGACCCCAACCCCCAGGAGGTGAACCTGACCAACGTG
ACCGAGAAGTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCATGAG
GACATCATCAGCCTGTGGGACCAAGCCTGAAGCCCTGCGTGGGCGCCGGCC
AGGCCTGCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCC
CGCCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGC
CCCTGCAAGAACGTGAGCACCGTGCAAGTGCAACCGCATCCGCCCCGTGG
TGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGAGATCGTGCT
GCGCTCCGAGAACTTACCCGACAACGCCAAGACCATCATCGTGACGTGAAC
GAGTCCGTGGAGATCAACTGCATCCGCCCAACAACAACACGCGTAAGAGCA
TCCACATCGGCCCGGCCGCGCTTCTACGCCACCGCGGACATCATCGCGCA
CATCCGCCAAGGCCCACTGCAACATCAGCAAGGCCAAGTGGACCAACACCCCTC
GAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGACCATC
ATCTTCAACAGCAGCAGCAGCGCGCGCATCCCGGAGATCGTGTTCACAGCTTCA
ACTGCGGCGCGAGTTCTTCTACTGCAACACCAAGCCAGCTGTTCACAGCAC
CTGGAACATCACCGAGGAGGTGAACAAGCAAGGAGAACGACACCATCAT
CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAGGAGGTGGGCAAG
GCCATGTACGCCCCCCCATCCGCGGCCAGATCAAGTGCAAGCAGCAATATTA
CCGGCCTGTGTGACCCCGGACCGCGCGGCAACATGAAGGACAACCGCACCAACGA
CACCGAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAACCGCACCAACGA
GAGCTGTACAAGTACAAGGTGGTGCAGCATCGAGCCCTTGGGCTGGCCCCCA
CCGAGGCCAAGCGCCGCGTGGTGACGCGCAGAGAAGCGCGCCCTGGGCTGG
GCGCCCTGTTTCATCGGCTTCTGGGCGCGCGCGGAGCACCATGGGCGCGCG
CTCCGTGACCTTGACCGTGCAAGCCCGCCAGCTGCTGAGCGGCATCGTGACG
CAGCAGAAACAACCTGCTGCGCGCCATCGAGGCCAGCAGCACCCTGTGACG
TGACCGTGTGGGGCATCAAGCAGCTGCAAGGCCCGCATCCTGGCCGTGGAGCG
CTACCTGAAGGACCAAGCAGCTGCTGGGCATCTGGGCTGCAAGCGGCAAGCTG
ATCTGCACCAACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGA
CCGAGATCTGGGACAACATGACCTGGATGGAATGGGAGCGCGAGATCGGCA
ACTACACCGGCTGATCTACAACCTGATCGAGATCGCCAGAACCAAGCAGGA
GAAAGACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAA
CTGGTTCGACATCAACCTGGCTGTGGTACATCTAAGATATCGGATCTCTTA
GA

FIG. 46
(SEQ ID NO:59)

Gp140modUS4.DV2

GAATTCCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGC
TGTTGGGAGCAGTCTTCGTTTCGCCACGCGCCACCACCGTGTGTGGGTGACC
GTGTACTACGGCGTGCCCGTGTGGAAAGGAGGCCACCACCCTGTCTGCG
CCAGCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTGGGCCACCCA
CGCCTGCGTGCACCGACCCCAACCCCAAGGAGGTGAACCTGACCAACGTG
ACCGAGAACTTCAACATGTGGAGAACAACATGGTGGAGCAGATGCATGAG
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCTGCGTGAAGCTGACCC
CCCTGTGCGTGACCCTGAACTGCACCGACAAGCTGACCGGCAGCACCAACGG
CACCAACAGCACCAGCGGCACCAACAGCACCAGCGGCACCAACAGCACCAG
CACCAACAGCACCAGACAGCTGGGAGAAGATGCCCGAGGGCGAGATCAAGAA
CTGCAGCTTCAACATCGGCGCCGGCCGCTGATCAACTGCAACACCAGCGTG
ATCACCCAGGCCTGCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACT
GCGCCCGCGCGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGG
CACCGGCCCTGCAAGAACGTGAGCACCCTGCACTGCAACCCAGCGCATCCG
CCCGTGGTGAGCACCCAGCTGCTGTGAACGGCAGCTGGCCGAGGAGGA
TCGTGCTGCGTCCGAGAACTTCAACGACAACGCCAAGACCATCATCGTGCA
GCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAAACAACGCGT
AAGAGCATCCACATCGGCCCGCGCGCGCTTCTACGCCACCGGCGACATCA
TCGGGCGACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAACTGGACCAA
CACCTCAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCCGGAACAACAA
GACCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGATCGTGTTCAC
AGCTTCAACTGCGCGCGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAA
CAGCACTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACAC
CATCATCTCGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAGGAGGTG
GGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCAAGCA
ATATTACCGGCTGCTGCTGACCCGCGACCGCGGCCAACAAACAACCCGAC
CAACGACACCGAGACCTTCCGCCCGCGCGGCCAACATGAAGGACAACCTG
GCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGAGCCCCCTGGGCGTG
GCCCCACCCAGGCCAAGCGCGCGGTGGTGACGCGGAGAAGCGCGCGCGTG
GGCCTGGGCGCCCTGTTTCATCGGCTTCTGGGCGCGCGGGAGCACCATGG
GCGCGCCCTCCGTGACCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCAT
CGTGACGACGACAGAACAACTGCTGCGCGCATCGAGGCCACGAGCACCTG
CTGCAGCTGACCGTGTGGGACATCAAGCAGCTGCAGGCCCGCATCTGCGCG
TGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCGTGCAGCGG
CAAGCTGATCTGCCACCAACCCGTGCCCTGGAACAGCAGCTGGAGCAACAAG
AGCCTGACCGAGATCTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAG
ATCGGCAACTACACCGGCTGATCTCAACCTGATCGAGATCGCCAGAACCC
AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCC
TGTGGAAGTGGTTCGACATCAACAACTGGCTGTGGTACATCTAAGATATCGG
ATCCTCTAGA

FIG. 47

(SEQ ID NO:60)

60 / 131

Gp140modmutUS4.DV1V2

GAATTCGCCACCATTGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGC
TGTGTGGAGCAGTCTTCGTTTCGCCCCAGCGCCACCACCGTGTCTGTGGGTGACC
GTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCACCCTGTTCTGCG
CCAGCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTGGGCCACCC
ACGCCTGCGTGCCACCACCCCAACCCCCAGGAGGTGAACCTGACCAACGT
GACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCATGA
GGACATCATCAGCCTGTGGACCAAGCCTGAAGCCCTGCGTGGGCCGCCGGC
CAGGCCTGCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCC
CCGCCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAGTTCAACGGGCACCCG
CCCTGCAAGAACGTGAGCACCCTGCACTGCAACCCACGGCATCCGCCCGCTG
GTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCGTGC
TGGCTCCGAGAACTTCAACGACAACGCCAAGACCATCATCGTGACGTGAA
CGAGTCCGTGGAGATCAATGCATCCGCCCCCAACAACAACCGCGTAAGAGC
ATCCACATCGGCCCGGCCGCGCCTTCTACGCCACCGGCGACATCATCGGCG
ACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCTT
CGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCCGGAACAACAAGACCAT
CATCTTCAACAGCAGCAGCGCGCGCGCACCCCGAGATCGTGTTCACAGCTTC
AACTGCGGCGGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAACAGCA
CCTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCA
TCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAGGAGGTGGGCAA
GGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGACGACGAGCAATATT
ACCGGCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCACCAACG
ACACCGAGACCTTCGCCCGCGCGCGGCAACATGAAGGACAACCTGGCGCA
GCGAGCTGTACAAGTACAAGGTGGTGCGCATCGAGCCCTGGGCGTGGCCCC
CACCCAGGCCAAGCGCCGCGTGGTGACGCGGAGAGAAGAGCGCCGTGGGCCT
GGGCGCCCTGTTATCGGCTTCTCTGGCGCGCGCGGAGCACCATGGGCGCC
GCCTCCGTGACCTGACCGTGACGCGCGCCAGCTGCTGAGCGGCATCGTGC
AGCAGCAGAAACAACCTGCTGCGCGCCATCGAGGCCCAGCAGCACTGCTGCA
GCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCCGCATCTGGCCGTGGAG
CGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGC
TGATCTGCACCACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCT
GACCGAGATCTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGG
CAACTACACCGGCTGATCTACAACCTGATCGAGATCGCCAGAACCAAGCAG
GAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGG
AACTGGTTCGACATACCAACTGGCTGTGTGATCTAAGATATCGGATCCTC
TAGA

FIG. 48

(SEQ ID NO:61)

gp140.mod.US4.del128-194

GAATTCCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGG
AGCAGTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTACTACGGCG
TGCCCCGTGTGGAAGGAGGCCACCAACACCTGTTCTGCGCCAGCGACGCCAAGGCTTAC
AAGGCCGAGGCCACAACTGTGTGGGCCACCCAGCGCTGCGTGCCCAACGACCCCAACCC
CCAGGAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGG
TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG
AAGCTGACCCCCCTGTGCTGGGGGAGGAACTGCGAGACCGAGCTGATCACCAGGC
CTGCCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTCG
CCATCCTGAAGTGCAAGGACAAGAGTTCAACGGCACCCGGCCCTGCAAGAAGCTGAGC
ACCGTGCAGTGCAACCCAGGCATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGG
CAGCCTGGCCGAGGAGAGATCGTGTGCGCTCCGAGAACTTCAACGACAACGCCAAGA
CCATCATCGTGCAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCCAACACAAC
ACGCGTAAGAGCATCCACATCGGCCCGCGCGCGCTTCTACGCCACCGGCAGATCAT
CGGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAACTGGAACCAACACCTCG
AGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGACCATCATCTTCAAC
AGCAGCAGCGCGCGGACCCCGAGATCGTGTTCACAGCTTCAACTGCGGCGCGAGTT
CTTCTACTGCAACACCGACCGAGCTGTTCAACAGCACCTGGAACATCACCGAGGAGGTGA
ACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCGGCATCCGCCAGATCATCAAC
ATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCAGATCAAGTG
CAGCAGCAATATTACCGCCCTGCTGCTGACCCGCGACGGCGGCACCAACAACGCCGA
CCAACGACACCGAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAACCTGGCGCAGC
GAGCTGTACAAGTACAGGTGTGTGCGCATCGAGCCCCCTGGGCGTGGCCCCACCCAGGC
CAAGCGCCGCGTGGTGAGCGCGAGAAGCGCCCGTGGGCTGGGCGCCCTGTTTCATCG
GCTTCTGGGCGCCCGCGGAGCACCATGGGCGCCGCTCCGTGAACCTGACCGTGCAG
GCCCGCAGCTGCTGAGCGGCATCGTGACGACGAGCAACACCTGCTGCGCGCCATCGA
GGCCCCAGCAGCACTGCTGACGTGACCGTGTGGGCATCAAGCAGCTGCAGGCCCGCA
TCTGCGCCCTGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGCAGC
GGCAAGCTGATCTGCACCAACCCGCTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCT
GACCGAGATCTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACA
CCGGCTGATCTACAACCTGATCGAGATCGCCAGAACAGCAGGAGAAGAAGCAGGAG
GAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAAGTGGTTCGACATCACCAACTG
GCTGTGGTACATCTAAGATATCGGATCCTCTAGA

FIG. 49
(SEQ ID NO:62)

62 / 131

gp140.mut.mod.US4.del128-194

GAATTCCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGG
AGCAGTCTTTCGTTTCGCCAGCGCCACCAACCGTGCTGTGGGTGACCGTGTACTACGCGC
TGCCCGTGTGGAGGAGGCCACCAACCCCTGTTCTGCGCCAGCGACGCCAAGGCTTAC
AAGGCCGAGGCCCAACAAGTGTGGGCCACCCACGCGCTGCGTGCCACCGACCCCAACCC
CCAGGAGGTGAACCTGACCAACGTGACCGAGAAGTTCAACATGTGGAAGAACAACATGG
TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG
AAGCTGACCCCCCTGTGCTGGGGGACGGGAAGTGCAGAGACCAGCGTGATCACCAGGC
CTGCCCCAAGGTGAGCTTCGAGCCCATCCCATCCACTACTGCGCCCCCGCGGCTTCG
CCATCCTGAAGTGCAGGACAAGAAGTTCAACGGCACCGGCCCTGCAAGAAGCTGAGC
ACCGTGCACTGCACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCTGTGCTGAACGG
CAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACTTACCAGACAACGCCAAGA
CCATCATCGTGCAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCAACAACAAC
ACGCGTAAGAGCATCCACATCGGCCCCGGCGCGCCTTACGCCACCGGCGACATCAT
CGGCGACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCCTCG
AGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGACCATCATCTTCAAC
AGCAGCAGCGCGGCGGCGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGGCGCGAGTT
CTTCTACTGCAACACAGCCAGCTGTTCAACAGCACCTGGAAACATCACCAGGAGGTGA
ACAAGACCAAGGAGAACGACACCATCATCTGCCCCGCGCATCCGCCAGATCATCAAC
ATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGCGGCCAGATCAAGTG
CAGCAGCAATATTACCGGCCTGTGCTGACCCGCGAGCGGCGCACCAACAACAACCGCA
CCAACGACACCGAGACCTTCCGCCCGGCGGCGGCAACATGAAGGACAACCTGGCGCAGC
GAGCTGTACAAGTACAAGGTGTGCGCATCGAGCCCCCTGGGCGTGGCCCCACCCAGGC
CAAGCGCCGCGTGGTGCAGCGCGAGAAGAGCGCCGTGGGCTGGGCGCCCTGTTTCATCG
GCTTCTTGGGCGCCGCGGGGAGCACCATGGGCGCGCCCTCCGTGACCTGACCCGTGCAG
GCCCGCCAGCTGCTGAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGGCCATCGA
GGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCA
TCCTGGCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGC
GGCAAGCTGATCTGCACCAACCAACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCT
GACCGAGATCTGGGACAACATGACCTGGATGGATGGGAGCGCGAGATCGGCAACTACA
CCGGCCCTGATCTACAACCTGATCGAGATCGCCAGAACCAAGCAGGAGAAGAACGAGCAG
GAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAAGTGGTTCGACATCACCACCTG
GCTGTGTACATCTAAGATATCGGATCCTCTAGA

FIG. 50

(SEQ ID NO:63)

gp160.modUS4

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA
GTCCTCGTTTGCGCCACGCCACCCGCTGCTGTGGGTGACCGGTGTACTACGGCGTGCCCGTG
TGGAAAGGAGGCCACCAACCCCTGTTCTGCGCCAGCAGCCCAAGGCTTACAAGGCCGAGGC
CCAACAACGTGTGGGCCACCCACGCTGCGTGCCACCGACCCCAACCCCAAGGAGGTGAACC
TGACCAACGTGACCCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCATGAG
GACATCATCAGCTGTGGGACCAAGCCTGAAGCCCTGCGTGAAGCTGACCCCTGTGCGTG
ACCTGAACTGCACCGACAAGCTGACCGGACGACCAACGGCACCAACAGCAGCAGCTGGGAGAAGATG
CAACAGCACCCGCGCACCAACAGCAGCAGCAACAGCAGCAGCAGCTGGGAGAAGATG
CCCCAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCAACGCGTGCGCGACAAGGTGCA
GAAGGAGTACAGCCTGTTCTACAAGCTGGAGCTGGTGGCCATCGACAACGACAACGCCAGCT
ACCGCTGATCAACTGCAACACCAAGCAGCTGATCACCAAGCCTGCCCAAGGTGAGCTTCGAGC
CCATCCCCATCCACTACTGCGCCCCCGCGGCTTCGGCATCCTGAAGTGCAAGGACAAGAAAT
TCAACGGCACCAGCCCTGCAAGAACGTGAGCAACCGTGCAAGTGCACCCAGCAGTCCGCCCC
GTGGTGAGCACCCAGCTGCTGCTGAAACGGCAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTC
CGAGAACTTCACCGACAACGCCAAGACCATCATCGTGCAAGTGAACGAGTCCGTGGAGATCA
ACTGCATCCGCCCAACAACAACAGCGTAAGAGCATCCACATCGGCCCGCGCGCGCTTCT
ACGCCACCGCGACATCATCGCGACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAAG
TGGACCAACACCCCTGAGCAGATCGTGGAGAAAGTGGCGGAGCAGTTCGGCAACAACAAGAC
CATCATCTTCAACAGCAGCAGCGCGCGGACCCCGAGATCGTGTCCAACGCTTCAACTGCGG
CGCGAGTCTTCTACTGCAACACCAAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGA
GGTGAAACAAGCAAGGAGAACGACACCATCATCTGCGCTGCGCATCCGCCAGATCATCA
ACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCGCCCATCCGCGGCCAGATCAAGTGC
AGCAGCAATATTACCGGCTGCTGCTGACCCGCGACGGCGGCAACAACAACCGCAACAA
CGACACCGAGACCTTCGCCCGCGCGCGCAACATGAAGGACAACCTGGCGCAGCGAGCTGT
ACAAGTACAAGGTGGTGGCATCGAGCCCTGGGCGTGGGCCCAACCCAGGCCAAGCGCGCG
GTGGTGACAGCGGAGAAAGCGCGCCGTGGGCTGGGCGCTGGGCCCAACCCAGGCCAAGCGCGCG
GCCGGAGCAACATGGCGCCGCTCCGTGACCTGACCGTGACGCGCCGCAAGCTGCTGAG
CGGCATCGTGACGACGACGAACAACCTGCTGCGCGCATCGAGGCCAGCAGCACTGCTGC
AGCTGACCGTGTGGGCATCAAGCAGCTGACAGGCCGCACTCTGGCGCTGGAGCGCTACCTG
AAGGACACGACAGCTGTGGGCATCTGGGGCTGCAAGCGCAAGCTGATCTGCACCAACCCGCT
GCCCTGGAACACGAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAACATGACCTGGA
TGGAGTGGGAGCGCGAGATCGGCAACTACCCGGCTGATCTACAACCTGATCGAGATCGCC
CAGAACACGAGGAGAAAGAACGAGCAGGAGCTGCTGAGCTGTAACCGCTGCGCCAGGGCT
GGAACTGGTTCGACATCAACAACTGGCTGTGGTACATCCGCATCTTCACTATGATCGTGGGCG
GCCTGATCGGCTGCGCATCGTGTTCGCGTGCTGAGCATCGTGAACCGCTGCGCCAGGGCT
ACGCCCCATCAGCCTGACAGACCCGCTGCGCGCGCGCGCAGCCGAGCAACCGCTGGTGACCGCTGCT
ATCGAGGAGGAGGCGCGCAGCGCGACCGCGACCGCAGCAACCGCTGGTGACCGCTGCTGCT
GGCCCTGATCTGGGAGCACTGCGCAGCTGTGCTGTTACGCTACCAACCGCTGGCGAGCT
GCTGCTGATCGTGGCCGCTGCTGGAGCTGCTGGGCGCGCGGCTGGGAGGCCCTGAAGT
ACTGGTGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGCAGCGCGCTGAGCTGTTT
AAGGCCACCGCCATCGCCGTGGCCAGGGCAACCGCCATCATCGAGTCTGTGACGCGCAT
CTTCGCGCGTGATCCATCCCCCGCGCATCCGCCAGGGCTGGAAGCGCGCCCTGCTGTA
AGATATCGGATCCTCTAGA

FIG. 51

(SEQ ID NO:64)

gp160.modUS4.delV1

GAA TTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA
GTCTTCGTTTCGCCACGCGCCACCAACCGTGCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGT
TGGAAGGAGGCCACCAACACCGCTGTCTGCGCCACGCGACGCCAAGGCTTACAAGGCCGAGGC
CCACAACGTGTGGGCCACCAACGCCTGCGTGCCACCGACCCCAACCCCGAGGAGGTGAAC
TGACCAACGTGACCGGAACTTCAACATGTGGAAAGAACATAGGTGGAGCAGATGCAATGAG
GACATCATCAGCCTGTGGGACAGACCTGAAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG
ACCTGAACTGCACCGACAAGCTGGGCGCGCGCGGAGATCAAGAACTGACAGCTTCAACAT
CACCAACGCGTGCAGCAAGGTGCAAGAGGAGTACAGCCTGTTCTACAAGCTGGACGTGG
TGCCCATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCAACACGCGTGATCACCC
AGGCCTGCCCAAGGTGAGCTTGAAGCCCATCCCATCCACTACTGCGCCCCGCGCGCTTGG
CCATCCTGAAGTGCAAGGACAAGAGTTCAACGGCACCGGCCCTGCAAGAACGTGAGCAAC
GTGCAGTGCACCACGGCATCCGCCCGTGGTGAGCACCCAGCTGCTGTAAGCGGACGCTG
GCGGAGGAGGATCGTGCTGCGCTCCGAGAACTTCAACGACAACGCCAAGACCATCATCTGT
GCAAGTGAAACGAGTCGTGGAGATCAACTGCATCCGCCCAACAACAACACGCGTAAGAGCA
TCCACATCGGCCCGCGCGCCTTCTACGCCACCGCGGACATCATCGGCGACATCCGCCAGG
CCCACTGCAACATCAGCAAGGCCAACTGGACCAACCCCTCGAGCAGATCGTGAGAAAGTGT
CGCGAGCAATTGGGCAACAACAAGACCATCATCTTCAACAGCAGCAGCGCGCGGACGCCGA
GATCGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTGCAACCAACACCGCAGCTGT
CAACAGCACTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCATCT
TGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCC
CCCCCATCCGCGGCGAGATCAAGTGACAGCAATATTACCGGCTGTGCTGTGACCCGCGAC
GGCGGCAACAACAACCGCACCAACGACACCGAGACCTTCCGCCCGCGCGGCGGCAACAT
GAAGGACAACTGGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGAGCCCTGGGG
TGCCCCCAACCCAGGCCAAGCGCGCGTGGTGACGCGGAGAAGCGCGCGCTGGGGCTGGG
GCCCTGTTCATCGGCTTCTTGGGCGCGCGCGGAGCACCATGGGCGCGCTCCGTGACCGCTG
ACCGTGACGCGCGGAGCTGTGAGCGGCATCGTGACGAGCAGCAACCTGCTGCGCGC
CATCGAGGCCAGCAGCACCTGCTGACGCTGACCGTGTGGGCGATCAAGCACTGACAGGCC
GCATCTGTGCGGTGGAGGCTACCTGAAAGGACCAAGCAGCTGCTGGGCACTTGTGGGCTG
GGCAAGCTGACTGCAACCAACCGTGCCTGGAACAGCAGCTGGAGCAACAGAGCCTGAC
CGAGATCTGGGCAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCACTACACCGGCC
TGATCTACAACCTGATCGAGATCGCCAGAACCCAGCAGGAGAAGAACGAGCAGGAGCTGCTG
GAGCTGGACAAGTGGGCCAGCTGTGGAATCGGACATCAACCACTGGCTGTGGTATCATC
CGCATCTTCATCATGATCGTGGGCGGCTGATCGGCTGCGCATCGTGTTCGCGCTGCTGAGC
ATCGTGAACCGGTGCGCAGGGGCTACAGCCCATCAGCTGCAACCGCCTGCCCGCCGAG
CGCGCGCGGACCGCGCCGAGGGCATCGAGGAGGAGGGCGCGAGCGGACCGCGACCGCA
GCAACCGCCTGGTGACGGCTGCTGGCCCTGATCTGGGACGACCTGCGCAGCCTGTGCTGT
TCAGATCAACACCGCTGCGGACCTGCTGCTGATCTGGGCCGATCTGGAGCTGCTGGGCG
GCGCGGCTGGGAGGCCCTGAAGTACTGGTGAACCTGTGCACTGAGGCCAGGAGCTG
AAGAGCAGCGCGGTGAGCCTGTTCAACGCCACCGCATCGCCGTGGCGGAGGCGCACCG
CATCATCGAGATCGTGACGCGCATCTCCGCGCGGTGATCCACATCCCGCGCGCATCCGCCA
GGGCTGGAGCGCGCCTGCTGTAAAGATATCGGATCTCTAGA

FIG. 52

(SEQ ID NO:65)

65 / 131

gp160.mod.US4.delV2

GAATTGCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGG
AGCAGTCTTTCGTTTCGCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTACTACGGCG
TGCCCGTGTGGAAGGAGGCCACCACCCTGTCTGCGCGCAGCGACGCCAAGGCTTAC
AAGGCCGAGGCCCAACAAGTGTGGGCCACCCACGCTGCGTGCCACCGACCCCAACCC
CCAGGAGGTGAACCTGCACAACTGACCGAGAATTCAACATGTGGAAGAACAACATG
TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACCAAGACCTGAAAGCCCTGCGGTG
AAGCTGACCCCTGTGCGTGACCTGAACTGCACCGACAAGCTGACCGGCAGACCAA
CGGCACCAACAGCACCAGCGGCACCAACAGCACCAGCGGCACCAACAGCACCAGCACA
ACAGCACCGACAGCTGGGAGAAGATGCCCCAGGGCGAGATCAAGAAGTGCAGCTTCAAC
ATCGCGCGCGCGCGCTGATCAACTGCAACACAGCGTGATCACCAGGCTGCCCAA
GGTGAGCTTCGAGCCATCCCATCCACTACTGCGCCCCCGCGGCTTCGCCCATCTGTA
AGTGCAAGGACAAGAACTTCAACGGCACCGGCCCTGCAAGAAGCTGAGGACCGTGAG
TGCAACCCACGGCATCCGCCCGCTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGC
CGAGGAGGAGATCGTGCTGCGCTCCGAGAATTACCGGACAACGCCAAGACCATCATCG
TGCAAGTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCAACAACAACAGCGTAAG
AGCATCCACATCGGCCCGCGCGCGCTTCTACGCCACCGGCAGCATCATCGGCGCAT
CCGCCAGGCCCACTGCAACATCAGCAAGGCCAATCGGACCAACACCTCGAGCAGATCG
TGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGACCATCATCTTCAACAGCAGCAGC
GGCGCGCAGCCCGAGATCGTGTTCCACAGCTTCAACTGCGCGCGCGAGTTCTTCTACTG
CAACACCAGCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAGGTGAACAAGACA
AGGAGAACGACACCATCATCTGCCCTGCGCGCATCCGCCAGATCATCAACATGTGGCAG
GAGGTGGGCAAGGCCATGTACGCCCGCCCATCCGCGGCAGATCAAGTGCGAGCAGCAA
TATTACCGGCTGCTGCTGACCCCGCGAGCGGCACCAACAACAACCGCACCAACGACA
CCGAGACCTTCGCCCGCGCGCGGCAACATGAAGGACAACCTGGCGCAGCGAGCTGTAC
AAGTACAAGGTGTTGCGCATCGAGCCCTGCGCGTGCGCCCCACCGAGGCCAAGCGCG
CGTGCTGACGCGCAGAGAGCGCGCGTGGGCTGGGCGCCTGTTTATCGGCTTCCTGG
GCGCGCGCGGAGCACCATGGGCGCGCCCTCGTGACCTGACCGTGACGCGCCGCGCAG
CTGCTGAGCGGCATCGTGACGACGACGAACAACCTGCTGCGCGCATCGAGGCCGCGCA
GCACCTGCTGAGCTGACCGTGTGGGTCATCAAGCAGCTGCAGGCCGCGCATCTGGCGG
TGGAGCGCTACCTGAAGGACACGAGCTGCTGGGTCATCTGGGCTGACGCGGCAAGCTG
ATCTGCACACACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCTGACCGAGAT
CTGGGACAACATGACCTGGATGGAGTGGGAGCGGAGATCGGCAACTACACCGGCTGA
TCTACAACCTGATCGAGATCGCCAGAACAGCAGGAGAAGAAGCAGCAGGAGCTGCTG
GAGCTGGACAAGTGGGCGAGCTGTGGAACCTGTTTCGACATCACCAACTGCGCTGTGGTA
CATCCGCTATCTTATCATGATCGTGCGCGCGCTGATCGGCGCATCAGCCTGCGAGACCCG
TGCTGAGCATCGTGAACCGCGTGCGCGAGGCTACAGCCCATCAGCCTGCGAGACCCG
CTGCCCGCCACGCGCGGCCGACCGCGCCGAGGGCATCGAGGAGGAGGGCGCGCGAGCG
CGACCGCGACCGCAGCAACCGCTGGTGACCGCGCTGCGCCCTGATCTGGGACGACC
TGCGAGCCTGTGCTGCTTTCAGCTACACCGCTGCGCGACCTGCTGCTGATCGTGCGC
CGCATCGTGGAGCTGCTGGGCGCGCGCGCTGGGAGGCCCTGAAGTATCTGGTGAACCT
GCTGCACTACTGGAGCCAGGAGCTGAAGAGCAGCGCGCTGAGCCTGTTTCAACGCCACCG
CCATCGCGTGCGCGAGGCGACCGACCGCATCATGAGATCGTGCGAGCGCATCTTCCG
GCCGTGATCCACATCCCCCGCGCATCGGCCAGGCGCTGGAGCGCGCCCTGCTGTGAAGA
TATCGATCTCTAGA

FIG. 53

(SEQ ID NO:66)

66 / 131

gp160.modUS4delV1/2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA
GTCTTCGTGTTTCGCCACGCCCAACACCGTGCTGTGGGTGACCGTGTACTACGGCGTGCCCGTG
TGGAAAGGAGGCCACCAACACCGTGTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC
CCACAACCGTGTGGGCCACCCACGCTGCGTGCCACCGACCCCAACCCCGAGGAGGTGAACC
TGAACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGTTGGAGCAGATGCATGAG
GACATCATCAGCCTGTGGGACCAGACCTGAAAGCCCTGCGTGGCGCCGCGCCAGGCTGCC
CAAGGTGAGCTTCGAGCCCATCCCATCCACTACTGCGCCCCCGCGGCTTGCCTCTGTAA
GTGCAAGGACAAAGAGTTCAACGGCAACCGGCCCTGCAAGAACGTGACACCGTGCAATGCA
CCCCGGCA TCCGCCCGTGGTGAGCAACCAAGTGTGCTGCTGAAACGGCA GCCTGGCCGAGGAG
GAGATCGTGTGCGCTCCGAGAACTTCAACGCAACCGCAAGACCATCATCGTGACGTGAA
CGAGTCCGTGGAGATCAACTGCATCCGCCCAACAACAACACGCGTAAGAGCATCCACATCG
GCCCCGGCCGCGCTTCTACGCCACCGGCCAGATCATCGGCCGACATCCGCCAGGCCCATCGCA
ACATCAGCAAGGCCAACTGGACCAACCCCTCGAGCAGATCGTGGAGAAGCTGCGCGAGCAG
TTCGGCAACAAACAAGACCATCATCTTCAACAGCAGCAGCGCGCGCGACCCCGAGATCGTGT
CCACAGCTTCAACTGCGCGCGCGAGTTCTTCTACTGCAACACCGCCAGCTGTTCAACAGCA
CTGGAAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCATCTGCCCTGCC
GCATCCGCCAGATCATCAACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCATC
CGCGGCCAGATCAAGTGCAAGCAGCAATAATTACCGCGCTGCTGCTGACCCCGCAGCGCGGCAC
CAACAACAACCGCACCAACGACACCGAGACTTCCGCCCGCGCGCGCAACATGAAGGACA
ACTGGCGCAGCGAGCTGTACAAGTACAAGTGAAGGTGGTGCGCATCGAGCCCTGGCGCTGGCC
ACCCAGGCCAAGCGCGCGTGGTGACGCGGAGAAGCGCGCGTGGCGCTGGGGCCCTGTT
CATCGGCTTCTGGCGCCGCCGGGAGCACCATGGCGCGCCCTCGTGACCTGACCGTGCA
GGCCCGCAGCTGCTGAGCGGCACTGTGACGACGACGAGAACAACTCTGCGCGCATCGAGG
CCCAAGCAGCACTGTGCAAGTACCGTGTGGGCACTCAAGCAGCTGACGGCCGATCCTG
GCCGTGGAGCGCTACTGAAAGGACCAAGCAGTGTGGGCACTGGGGCTGCAAGCGCAAGCT
GATCTGCACCAACCGTGCCTGGAACAGCAGCTGGAGCAACAAGACGCTGACCGAGATCT
GGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACACCGCGCTGATCTAC
AACTGATCGAGATCGCCAGAACACAGCAGGAGAAGAACAGCAGGAGCTGTGGAGCTGG
ACAACTGGGCCAGCTGTGGAACCTGGTTCGACATCAACAACTGGCTGTGATACCCGATCT
TCATCATGATCGTGGCGCGCTGATCGGCTGCGCATCGTGTTCGCGTGTGAGCATCTGTGA
ACCGCGTGCAGCAGGCTACAGCCCATCAGCTGACAGACCGCTGCCCGCCACGCGCGC
CCGACCGCGCCGAGGGCACTGAGGAGGAGGCGCGGAGCGCGACCGGACCGCAGCAAC
GCTGTGTCAGCGCTGTGCGCTGATCTGGGACGACCTGCGCAGCTGTGCTGTTCAGCT
ACCAACGCTGCGCGACCTGCTGCTGATCGTGGCCGATCGTGGAGCTGTGGGCGCGCG
CTGGGAGGCGCTGAAGTACTGGTGGAACCTGCTGAGTACTGGAAGCAGGAGCTGAAGAGC
AGCGCGTGAGGCTGTCAACGCCACCGCATCGCGTGGCGGAGGGCAACGACCGCATCAT
GAGATCGTGAAGCGCATCTCCGCGCGTATCCATCCCCCGCATCCGCCAGGCGCTG
GAGCGCGCCTGCTGAAGATATCGGATCCTCTAGA

FIG. 54
(SEQ ID NO:67)

gp160.modUS4 del 128-194

GAATTCCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGTGTGGAGCA
GTC TTCGTTTCGGCCAGCGCCACCAACCGTGTGTGGGTGACCGTGTACTACGGCGTGCCCGTG
TGGAAGGAGGCCACCAACCAACCTGTTCTGCGCCACGCAACGCAAGGCTTACAAGGCCGAGGC
CCACAAACGTGTGGGCCACCAACCGTGTGCTGCCACCGACCCCAACCCCAAGGAGGTGAAC
TGACCAACGTGACGAGAACTTCAACATGTGGAAGAAACAATGGTGGAGCATGCATGAG
GACATCATCAACCTGTGGGACCAAGCCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGGTG
GGGGCAAGGAACTGCGAGACCAAGCGTGATCACCAAGCCCTGCCCAAGGTGAGCTTCGAGCC
CATCCCAATCCACTACTGCGCCCCGCGCGCTTCCGCAATCTGAAAGTGAAGGACAAGAAATT
CAACGGCACCGGCCCTGCAAGAACGTGAGCAACCGTGCAAGTGCACCCACGGCATCCGCCCG
TGGTGAGCAACCAAGTGTGCTGAACGCGAGCTGCGCGAGCAITTCGGCAACAAACAGGACC
GAGAACTTCAACGACAACGCCAAGACCATCATCGTGCAAGTGAACGAGTCCGTGGAGATCAA
CTGCATCCGCCCCAAACAACAACGCGTAAGAGCATCCACATCGGCCCGCGCGCTTCTA
CGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCCATGCAACATCAGCAAGGCCAACT
GGACCAACACCTCGAGCAGATCGTGGAGAAAGTGTGCGGAGCAITTCGGCAACAAACAGACC
ATCATCTTCAACAGCAGCAGCGCGCGACCCCGAGATCGTGTTCACAGCTTCAACTGCGGCG
GGCGAGTCTTCTACTGCAACACCAAGCCAGCTGTTCAACAGCACTGGAACATCACCGAGGAG
GTGAACAAGACCAAGGAAACGACACCATCATCTGCTGCCCTGCCGATCCGCCAGATCATCAA
CATGTGGCAGGAGGTGGCAAGGCCATGTAGCCGCCCTGCGCGATCCGCCAGCAACAAAGACC
GCAGCAATAATTAACGGCTGTGCTGTAACCGCGAGCTGCGCGGACCAACAACACCGCAACAA
GACACCGAGACTTCCGCCCGCGCGCGGCAACATGAAGGACAACTGGCGCAGCGAGCTGTA
CAAGTACAAAGTGTGCGCATCGAGCCCTGGGCGTGGCCCCACCAACGGCAAGCGCGCG
TGGTGACGCGGAGAAAGCGCGCGTGGGCTGGGCGCCCTGTTATCGGCTTCTGGGCGCGG
CCGGGAGCAACATGGGCGCGCGCTCGGTGACCTGACCGTGACGGCCGCCAGCTGCTGAGC
GGCATCGTGACGACGAGAAACAACCTGTGCGCGCATCGAGGCCCAAGCAGCACCTGCTGCA
GCTGACCGTGTGGGCGCATCAAGCAGTGCAAGGCCGATCTGGGCAACATGACCTGGAT
AGGACCAAGCAGCTGCTGGGCTGCAAGCGGCAAGCTGATCTGCAACCAACCGGTG
CCCTGGAACAGCAGCTGGAGCAACAGAGCTGACCGAGATCTGGGCAACATGACCTGGAT
GGAATGGGAGCGCGAGATCGGCAACTACACCGGCTGATCTACAACCTGATCGAGATCGCCC
AGAACAACGAGAGAGAAACGAGCAGGAGCTGTGGAGCTGGACAAGTGGGCCAGCCTGTG
GAACTGGTTCGACATCACCAACTGGCTGTGGTACATCCGATCTTCAATCATGTGTGGGCGG
CCTGATCGGCTGCGCATCGTGTTCGCGTGTGAGCATCGTGAACCGCGTGGCGCAGGGGCTA
CAGCCCCATCAGCTGACAGACCCGCTGCCGCCAGCGCGGCCCGGACCGCCCGAGGCA
TCGAGGAGGAGGGCGGAGCGGACCGGACCGGACCAACCGCTGTGCAAGCGCTGTGCTG
GCCCTGATCTGGGAGACCTGCGCAGCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
CTGTGATCGTGGGCCGATCGTGGAGCTGCTGGGCGCGCGGCTGGGAGGCCCTGAAAGTAC
TGGTGGAACTGCTGCGATCTGGAGCCAGGAGCTGAAGAGCAGCGCGTGGAGCTGTTCAA
CGCCACCGCCATCGCGGTGGCGGAGGACCGGACCGCATCATCGAGATCGTGGCAGCATCTT
CCGCGCGTGTACCATCCCCGCGCATCGGCCAGGGCTGGAGCGCGCCTGTGTGAAGA
TATCGGATCCTTAGA

FIG. 55
(SEQ ID NO:68)

68 / 131

Env_US4_C4wt

GACACTATCATACTCCCATGCAGAATAAGACAAATTATAAACATGTGGCAAGAAGTAGG
AAAAGCAATGTATGCCCCCTCCCATCAGAGGACAAATTAAATGTTTCATCAAATATTACAG
GGCTGCTATTAACTAGAGATGGTGGT

FIG. 56

(SEQ ID NO:69)

Env_SF162_C4wt

GGAACTATCACACTCCCATGCAGAATAAAACAAATTATAAACAGGTGGCAGGAAGTAGG
AAAAGCAATGTATGCCCCCTCCCATCAGAGGACAAATTAGATGCTCATCAAATATTACAG
GACTGCTATTAAACAGAGATGGTGGT

FIG. 57

(SEQ ID NO:70)

Env_US4_C4mod

GACACCATCATCCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAGGAGGTGGG
CAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCAGCAGCAACATCACCG
GCCTGCTGCTGACCCGCGACGGCGGGC

FIG. 58

(SEQ ID NO:71)

Env_SF162_C4mod

GGCACCATCACCTGCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCAGGAGGTGGG
CAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCCGCTGCAGCAGCAACATCACCG
GCCTGCTGCTGACCCGCGACGGCGGGC

FIG. 59

(SEQ ID NO:72)

69 / 131

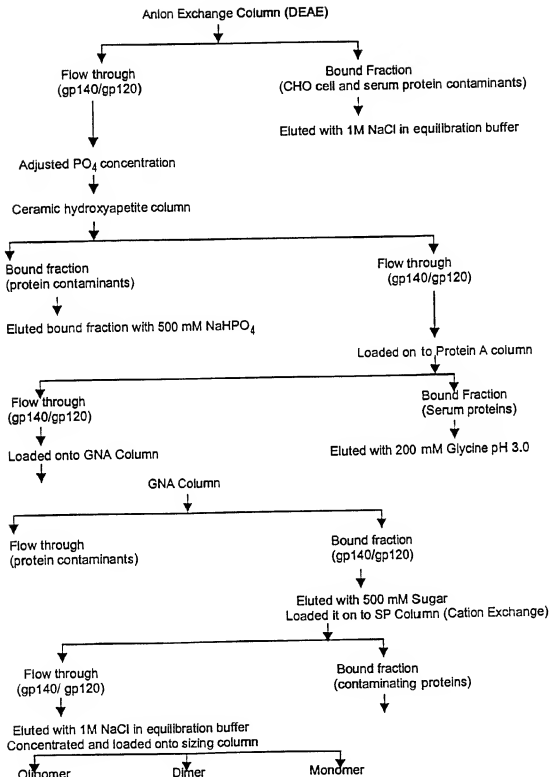


FIG. 60

gp160mod.us4.gag.modSF2

GAAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGA
 GCAGTCTTTCGTTTCCGCCAGCGCCACCCACCGCTGTGTGGGTGACCGTGTACTACGGCGTG
 CCCGTGTGGAAGGAGGCCACCCACCCCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAG
 GCCGAGGCCCAACAGTGTGGGCCACCCAGCGCTGCGTGCCACCGACCCCAACCCCCAG
 GAGGTGAACCTGACCAACCTGACCGGAACTTCAACATGTGGAAGAAACAACATGGTGGAG
 CAGATGCATGAGGACATCATCAGCCTGTGGGACGAGAGCCTGAAGCCCTGCGTGAAAGCTG
 ACCCCCCGTGTGCTGACCTGAACTGCACCGACAAGCTGACCGGAGCACCACCGGCACC
 GACAGCTGGGAGAAAGATGCCCGAGGGCGAGATCAAGAATGCAAGCTTCAACATCAACCACC
 AGCGTGCAGCAAGGTGACAGAGGAGTACAGCCTGTTCTACAAGCTGGAGCGTGGTGCCC
 ATCGACAACCAACGCCAGCTACCGCTGTATCAACTGCAACACCGAGCTGATCAACCCAG
 GCCTGCCCCCAAGGTGAGCTTCGAGCCCATCCCATCCACTACTGCGCCCCCGCGGCTTC
 GCCATCTGTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGCCCCCTGCAAGAAGCTGAGC
 ACCGTGAGCTGCACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCG
 AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACTTCAACGACAACGCCAAGACC
 ATCATCTGTGAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCAACAACAACACG
 CGTAAGAGCATCCACATCGGCCCGGCGCGCCTTCTACGCCACCGGCGACATCATCTGCG
 GACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCCTCAGACAG
 ATCGTGGAGAAAGCTGCGCGAGCAGTTTCGGCAACAACAAGACCATCTTCAACAGCAGC
 AGCGCGCGCGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGGCGCGAGTCTTCTTAC
 TGCAACACCGAGCCAGCTGTTCAACAGCACTGGAACATCACCAGGAGGAGTGAAACAAGACC
 AAGGAGAACGACACCATCATCTCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG
 GAGGTGGGCAAGGCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCAGCAGCAAT
 ATTACCGGCTGCTGCTGACCCGCGAGCGCGGCAACAACAACAACCGCACCAACGACCC
 GAGACCTTCCGCCCCCGCGCGGCAACATGAAGACAACCTGGCGCAGCGAGCTGTACAAG
 TACAAGGTGGTGCGCATCGAGCCCTGGGCGTGGCCCCACCAGGCCAAGCGCGCGTG
 GTGCAAGCGCAGAGAGCGCGCGTGGGCTGGGCGCCTGTTTATCGGCTTCTTGGGCGCG
 GCCGGGAGCACCATGGGCGCGCGCTTCGTGACCTGACCGTGAGGCCCGCGAGCTGTGCTG
 AGCGGCATCGTGACGAGCAGAAACCTGCTGCGGCCATCAGAGCCCGAGCAGCACTGTG
 CTGCAGCTGACCGCTGTGGGGCATCAAGCAGCTGCAGGCCCGCATCTTGGCGGTGGAGCGC
 TACCTTGAAGGACCGAGCTGCTGGGCATCTGGGGTGCAGCGGCAAGCTGATCTGCACCC
 ACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGACCTGACCGAGATCTGGGCAAC
 ATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCTGTATCTACAACCTG
 ATCGAGATCGCCAGAACACGACGAGGAGAAGACGAGCAGGAGCTGCTGGAGCTGGACAAG
 TGGGCCAGCCTGTGGAACCTGTTTCGACATCAACCACTGGCTGTGTTACATTCGCGATCTTC
 ATCATGATCGTGGGCGGCTGATCGGCCGCGCATCGTGTTCGCCGTGCTGAGCATCGTG
 AACCGGTGCGCCAGGGCTACAGCCCATCAGCCTGCAGACCCGCGCTGCCCGCCAGGGC
 GGCCTCCGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGCGAGCGCAGCCGACCCGAGC
 AACCCGCTGGTGACAGCGCTGCTGGCCTGATCTGGGACGACCTGCGCAGCCTGTGCTCTG
 TTACGTACACCGCTGCGCGACCTGCTGCTGATCTGTGGCCGACATCGTGGAGCTGCTG
 GGCCTGCGCGGTGGGAGGCCCTGAAGTACTGGTGAACCTGCTGCAGTACTGGAGCCAG
 GAGCTGAAGAGCAGCGCGGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGAGCCGAGGGC
 ACCGACCGCATCTGAGATCGTGACGCGCATCTTCGCGCGCTGTATCCACATCCCCCGC
 CGCATCCGCCAGGGCTGGAGCGCGCCTGCTGTAAGATATCGGATCCCTAGAGAATTC

FIG. 61A

(SEQ ID NO:73)

71 / 131

CGCCCCCCCCCCCCCCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGG
TTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTCCACCATATTGCCGCTTTT
GGCAATGTGAGGGCCCGGAAACCTGGCCCTGCTCTTGTACGAGCATTCCTAGGGGCTT
TCCCCCTCTGCCAAAGGAATGCAAGGTCTGTTGAATGTGCTGAAGGAAGCAGTTCCTCTG
GAAGCTTCTTGAAGCAAAACAACGCTGTAGCGACCCCTTTGCAGGCAGCGGAACCCCA
CCTGGCGCAGGTGCGCTCTGCGGCCAAAGCCACGTGTATAAGATACACCTGCAAAAGGCG
GCACAACCCCACTGCCACGTTGTGAGTTGGATAGTTGTGGAAGAGTCAAATGGCTCTCC
TCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAAGGTACCCCATTTGATGGGATCT
GATCTGGGGCCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAACGCTCTA
GGCCCCCCGAACACGCGGACGTGGTTTTCTTTGAAAAACACGATAATACCATTGGGCGC
CCGCGCCAGCGTCTGAGCGGCGGCGAGCTGGACAAGTGGGAGAAGATCCGCTGCGCCC
CGCGCGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGGGCCAGCCGCGAGCTGGAGCG
CTTCGCCGTGAACCCCGCCTGCTGGAGACCAGCGAGGGCTGCCGCCAGATCCTGGGCCA
GCTGCAGCCCGAGCTGCAGACCGGCAGCGAGGAGCTGCGCAGCCTGTACAACACCGTGGC
CACCTCTGACTGCTGTCACCGCGCATCGACGTCAAGGACACCAAGGAGGCCCTGGAGAA
GATCGAGGAGGAGCAGAACAAAGTCCAAAGAAGAGGCCAGCAGGCCGCGCGCCGCCCGG
CACCGGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATCTGTCAGAACTGCAAGGCCCA
GATGGTGCAACAGGCCATCAGCCCCGACCCCTGAACGCTTGGTGAAGGTGGTGGAGGA
GAAGGCCTTCAGCCCGAGGTGATCCCCATGTTGACGCCCTGAGCGAGGGGCCCAACCC
CCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCCACAGGCCGCTGCACGCCCGG
GAAGGAGACCATCAACGAGGAGGCCGCCGAGTGGGACCCGCTGCACCCCTGCACGCCCG
CCCCATCGCCCCCGGCAGATGCGCGAGCCCCGCGGCAGCGACATCGCCGGCACCACAG
CACCTTCAGGAGCAGATCGGCTGGATGACCAACACCCCCCATCCCCGTGGCGGAGAT
CTACAAGCGGTGGATCATCTGGGCCTGAACAAGATCGTGGCGGATGTACAGCCCCACAG
CATCTGGACATCCGCGAGGGCCCCAAGGAGCCCTTCGCGACTACGTGGACCGCTTCTA
CAAGACCTTGC GCGCTGAGCAGGCCAGCCAGGACGTGAAGAACTGGATGACCGAGACCT
GCTGGTGAGAACCGCCAAACCCGACTGCAAGACCATCTGAAGGCTCTCGGCCCGCGCGG
CACCTTGAGGAGATGATGACCGCCTGCCAGGGCGTGGGGCGGCCCGGCCCAAGGCCCG
CGTGCTGGCCGAGCGGATGAGCCAGGTGACGAACCCGCGGACCATCATGATGACGCGCG
CAACTTCGCAACAGCGGAAGACCGTCAAGTGCTTCAACTGCGGCAAGGAGGGCCACAC
CGCCAGGAAGTSCCGCGCCCCCGCAAGAAGGGCTGCTGGCGCTGCGGCCGCGAGGGCCA
CCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCCTGGGCAAGATCTGGCCAGCTA
CAGGGGCGCCCCCGGCAACTTCCTGCAAGGCCGCCGAGCCACCGCGCCCCCGGAGGA
GAGCTTCCGCTTGGCGAGGAGAAGACACCCCGCAGCAGAAGCAGGAGCCCATCGACAA
GGAGCTGTACCCCTGACCAGCCTGCGCAGCCTGTCGGCAACGACCCAGCAGCCAGTA
AGAATTACAGACTCGAGCAAGTCTAGA

FIG. 61B
(SEQ ID NO:73)

72 / 131

gp160mod.SF162.gag.modSF2

GAATTGCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGG
AGCAGTCTTCGTTTCGCCCAGCGCGTGGAGAAGCTGTGGGTGACCGTGTA CTACGGCG
TGCCCGTGTGGAAGGAGGCCACCAACACCTGTTCTGCGCCAGCGACGCCAAGGCCTAC
GACACCGAGGTGCACAACGTGTGGGCCACCCACGCTGCGTGCCCAACCGACCCCAACCC
CCAGGAGATCGTGCTGGAGAAGCTGACCGAGA ACTTCAACATGTGGAGAACAACATGG
TGGAGCAGATGCACGAGGACATCATCAGCCTGTGGGACGAGCGCTGAAGCCCTGCGTG
AAGCTGACCCCCCTGTGCGTGACCTGCACTGCACCAACCTGAAGAACGCCACCAACAC
CAAGAGCAGCAACTGGAAGGAGATGGACCGCGCGAGATCAAGAACTGCAGCTTCAAGG
TGACCAACAGCATCCGCAACAAGATGCGAAGGAGTACGCCCTGTTCTACAAGCTGGAG
GTGGTGCCCATCGACAACGACAACACCAAGCTACAAGCTGATCAACTGCAACACCAAGCGT
GATCACCCAGGCGCTGCCCAAGGTGAGCTTCGAGCCCATCCCATCCACTACTGCGCCC
CCGCGCGCTTCGCCATCCTGAAGTGCAACGACAAGAAGTTCAACGGCAGCGGCCCTGCG
ACCAACGTGAGCACCGTGCAGTGCACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCT
GCTGCTGAACGGCAGCTGCGCCGAGGAGGGCGTGGTGATCCGCGAGCGAGA ACTTCAACG
ACAACGCCAAGACCATCATCGTGCACTGAAGGAGAGCGTGAGATCAACTGCACCCGCG
CCCAACAACACACCCCGCAAGAGCATCAACATCGGCCCCGCGCGCCCTTCTACGCCAC
CGGCGACATCATCGCGCATCCGCGAGGCCCACTGCAACATCAGCGCGGAGAAGTGGGA
ACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGCCAGTTGCGCAACAAGACCATC
GTGTTCAAGCAGAGCAGCGCGCGCGAGACCCGAGATCGTGATGCACAGCTTCAACTGCGG
CGGCGAGTCTTCTACTGCAACAGCACCCAGCTGTTCACAGCAGCTGGAAACAACACCA
TCGGCCCCAACAACACCAACGGCACCATCACCTGCCCTGCCGCA TCAAGCAGATCATC
AACCGCTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCCATCCGCGCCAGATCCG
CTGCAGCAGCAACATCACCGGCCCTGCTGCTGACCCCGCAGCGCGGCAAGGAGATCAGCA
ACACCCAGGAGATCTTCCGCCCGCGCGGCACATGCGCGACAACACTGCGCAGCGAG
CTGTACAAGTACAAGGTGGTGAAGATCGAGCCCTGGGCGTGCCCCCAACAAGGCCAA
GCGCGCGTGGTGAGCGCGGAGAACCATGGGCGCCGCGAGCCTGACCCGTGACCTGCAAGGC
CGCCAGCTGCTGAGCGGCATCGTGACGAGCAGAACAACTGCTGCGCGCATCGAGGGC
CCAGCAGCACTGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCGTGC
TGGCCGTGAGCGCTACTGAAAGGACAGCAGCTGCTGGGCATCTGGGGCTGCAGCGCG
AAGCTGATCTGCACCAACCGCGTGCCTGGAACGCCAGCTGGAGCAACAAGAGCTTGA
CCAGATCTGGAACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGACAACATACCA
ACCTGATCTACACCTGATCGAGGAGAGCCAGAACAGCAGGAGAAGAACCGAGCAGGAG
CTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAAC TGGTTCGACATCAGCAAGTGGCT
GTGGTACATCAAGATCTTATCATGATGCTGTGGGCGGCCCTGGTGGCCCTGCGCATCTGT
TCACTGCTGAGCATCGTGAACCGCGTGCGCGCAGGGCTACAGCCCCCTGAGCTTCGAG
ACCCGCTTCCCCGCCCCCCCGCGGCCCGACCGCCCGAGGGCATCGAGGAGGAGGGCG
CGAGCGCAGCGCGACCGCAGCAGCCCCCTGTGTCACGGCTGCTGGCCCTGATCTGGG
AGCAGCTGCGCAGCCTGTGCTGTTCAGCTACACCGCCTGCGCGACCTGATCTCTGATC
GCCGCCCGCATCGTGGAGCTGCTGGGCGCGCGCGCTGGGAGGCCCTGAAGTACTGGG
CAACTGCTGCAGTACTGATCCAGGAGCTGAAGAACAGCGCGCTGAGCCTGTTTCGAGC
CCATCGCCATCGCCCTGGGCCGAGGGCACCGACCGCATATCGAGGTGGGCCGCGCATC
GGCGCGCCTTCTGACATCCCCGCGCATCGCGCAGGGCTTCGAGCGCGCCCTGCT

FIG. 62A

(SEQ ID NO:74)

73 / 131

GTAAC TCGAGCAAGTCTAGAGAATTCCGCCCCCCCCCCCCCCCCCTCTCCCTCCCCC
CCCCCTAACGTTACTGGCCGAAGCCGCTTGGAAATAGGCCGGTGTGGCTTTGTCTATAT
GTTATTTCCACCATATTGCCGCTCTTTTGGCAATGTGAGGCCCGGAAACCTGGCCCTG
TCTTCTTGACGAGCATTTCTAGGGGTCTTCCCCCTCTCGCCAAAGGAATGCAAGGTCTG
TTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGACAACAACGCTCTGT
AGCGACCCCTTGCAGGCAGCGGAACCCCCACCTGGCGACAGGTGCCTCTCGGCCCAAA
AGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCCAGTGCCACGTTGTGAGT
TGGATAGTTGTGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGCTGAA
GGATGCCAGAAAGGTACCCATTGTATGGGATCTGATCTGGGGCTCGGTGCATGTCT
TTACATGTGTTTAGTCGAGGTTAAAAAACGCTTAGGCCCCCCGGAACCCAGGGGACGTG
GTTTTCTTTGAAAAACACGATAATACCATGGGCGCCCGCCAGCGTGTGAGCGGGC
CGAGGCTGGACAAGTGGGAGAAGATCCGCTGCGCCCCGCGCGCAAGAAGAAGTACAAG
CTGAAGCACATCGTGTGGGCCAGCCGCGAGCTGGAGCGCTTCGCCGTGAACCCCGGCT
CTGGAGACCCAGCGAGGGCTGCCGCCAGATCTTGGGCCAGCTGCAGCCCGAGCTTCGAGA
CCGGCAGCGAGGAGCTGCCGAGCTGTACAACCCGTGGCCACCTGTACTGCGTGCAC
CAGCGCATCGACGTCAAGGACACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAA
CAAGTCCAAGAAGAAGGCCAGCAGGCCGCGCCGCCCGCCGACCCGCAACAGCAGCC
AGGTGAGCCAGAACTACCCCATCGTGCAAGCTGCAGGGCCAGATGTTGCACAGGCC
ATCAGCCCCCGCACCTTGAACGCTGGGTGAAGGTGGTGGAGGAGAAGGCCCTCAGCCCC
CAGGCTGTATCCCATGTTTCCGCGCCTGAGCGAGGGGCCACCCCCCAGGACCTGAACA
CGATGTTGAACACCGTGGGCGGCCACAGGCCGCCATGCAGATGCTGAAGGAGACCATC
AACGAGGAGGGCCGCGAGTGGGACCGCTGCACCCCGTGCACGCCGCGGCCACCCCTG
CGGCCAGATGCGCGAGGCCCGCGCGCAGCGACATCGCCGCGCACACCCAGCACCCTG
AGCAGATCGGCTGGATGACCAACAACCCCCCATCCCCGTGGGCGAGATCTACAAGCG
TGGATCATCTGGGCTGAACAAGATCGTGGGATGTACAGCCCCACCGATCTCTGGA
CATCCGCCAGGGGCCCAAGGAGCCCTTCCGCGACTACGTGGACCGCTTCTACAAGACC
TGCGCGCTGAGCAGGCCAGCCAGGACGTGAAGAAGTGGATGACCGAGACCCCTGCTGGT
GGAGGAGATGATGACCCGCTTCCAGGGCGTGGGCGGCCCGGCCAAGGCCCGCGTGC
TGGCCGAGGCGATGAGCCAGGTGACGAACCCGCGACCATCATGATGCAGCGCGCAAC
TTCCGCAACCAAGCGGAAGACCGTCAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCG
CAGGAAGTCCGCGCGCCCCCGCAAGAAGGGCTGCTGGCGCTGCGGCCGCGAGGGCCACC
AGATGAAGGAGCTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCAGCTAC
AAGGCCCGCCCGGCAACTTCTGCGAGCGCGCCCGAGCCACCCGCCCCCGGAGGA
GAGCTTCCGCTTGGCGAGGAGAAGACCACCCAGCCAGAAAGCAGGAGCCCATCGACA
AGGAGCTGTACCCCTGACAGCGCTGCGCAGCTGTTCCGGCAACGACCCAGCAGCCAG
TAAGAATTACAGCTCGAGCAAGTCTAGA

FIG. 62B

(SEQ ID NO:74)

gp160modUS4.delV1/V2.gag.modSF2

GAATTCCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGA
 GCAGTCTTCGTTTCGCCAGCGCCACCAACCGTGCTGTGGGTGACCGTGTACTIONACGCGGTG
 CCCGCTGTGGAAGGAGGCCACCAACCAACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAG
 GCCGAGGCCCAACAGTGTGGGCCACCAACGCTGCGTGCCACCGACCCCAACCCCCAG
 GAGGTGAACCTGACCAACGTGACCGGAGAATTTCAACATGTGGAAGAACAACATGGTGGAG
 CAGATGCATGAGGACATCATCAGCCTGTGGGACCAAGCCCTGAAGCCCTCGCTGGGGCGCC
 GGCCAGGCCTGCCCAAGSTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCGCC
 GGCTTCGCCCATCTGAAGTGAAGGACAAGAAGTTCAACGGCACCGGCCCCCTGCAAGAAC
 GTGAGCACCGTGCAGTGCAACCAACGCGCATCCGCCCCCGTGGTGAGCAACCGAGTCTGCTG
 AACCGCAGCCTGGCCGAGGAGAGATCGTGCTGCGCTCCGAGAATTTACCGCAACGCC
 AAGACCATCATCGTGCACTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCAACCAAC
 AACACCGTAAGAGCATCCACATCGGCCCGGCCGCGCTTCTACGCCACCGCGACATC
 ATCGCGGACATCCGCCAGGCCCATGCAACATCAGCAAGGCCAACTGGACCAACACCCCTG
 GAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCCGCAACAACAAGACCATCATCTTCAAC
 AGCAGCAGCGCGCGGCAACCCGAGATCGTGTTCACAGCTTCAACTGCGCGCGGAGTTT
 TTCTACTGCAACACCAAGCCAGTGTTCACAGCAGCCTGGAACATACCGAGGAGGTGAAC
 AAGACCAAGGAGAAGCAGACCATCATCTGCGCTGCGCATCCGCCAGATCATCAACATG
 TGGCAGGAGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCAGC
 AGCAATATTACCGGCTGCTGCTGACCCGCGCAGCGCGGCAACCAACAACCGCACCACAA
 GACACCGAGACCTTCCGCCCGCGCGCGGCAACATGAAGGCAACTGGCGCAGCAGCTG
 TACAAGTACAAGGTGGTGGCATCGAGCCCTGGGCGTGGCCCCCACCCAGGCCAAGCGC
 CGCGTGGTGAGCGCGGAGAAGCGCGCCGTGGGCTGGGCGCCCTGTTTATCGGCTTCTCTG
 GCGCCGCGCGGAGACCATGGGCGCGCCCTCGTGACCTGACCGTGAGCGCCCGCCAG
 CTGCTGAGCGGCATCGTGCAAGCAGCAGAACCAACTGCTGCGCGCCATCGAGGCCACAGCAG
 CACCTGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCATCTCTGGCCGTG
 GAGCGCTACCTGAAGACCAAGCAGCTGCTGGGCATCTGGGCTGCAGCGGCAAGCTGATC
 TGCAACCAACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGACCTGACCGAGATCTGG
 GACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCTGATCTAC
 AACCTGATCGAGATCGCCCAAGAACAGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTG
 GACAAGTGGGCCAGCCTGTGGAATCGGTTCGACATCAACCAACTGGCTGTGGTATCATCCGC
 ATCTTTCATCATGATCGTGGGCGGCTGATCGGCTGCGCATCGTGTTCGCGTCTGAGC
 ATCGTGAACCGCGTGCGCCAGGGCTACAGCCCATCAGCCTGACAGCCCGCTGCGCGCC
 CAGCGCGGCCCGGACCGCCCCGAGGGCATCGAGGAGGAGGGCGCGCGAGCGGACCGCGAC
 CGCAGCAACCGCTGTGTGACCGGCTGCTGCGCTGATGTGGAGCGACCTGCGCAGCCTG
 TGCCCTGTTCAGCTACCAACCGCTGCGCGACCTGCTGATGATCGTGGCCCCGCATCGTGGAG
 CTGCTGGGCGCGCGGCTGGGAGGCCCTGAAGTACTGGTGGAACTGCTGCAGTACTGG
 AGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTTCAACGCCACCGCATCGGCTGGGCC
 GAGGGCAACGACCGCATCATCGAGATCGTGACGCGCATCTTCCGCGCGTGATCCACATC
 CCCGCGCATCGCGCAGGGCTGGAGCGCGCCCTGCTGTGAAGATATCGGATCTCTAGA
 GAATTCGCCCCCCCCCCCCCCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGA
 AGCCGCTTGAATAAGGCGGTGTGCGTTGTCTATATGTTATTTCCACCATATTGCCG
 TCTTTTGGCAATGTGAGGGCCCGGAAACCTGCGCTGCTCTTCTGACGAGCATTCCTAGG
 GGTCTTTCCCTCTGCCAAAGGAATGCAAGGTCTGTTGAATGTGCTGAAGGAAGCAGTT

FIG. 63A

(SEQ ID NO:75)

75 / 131

CCTCTGGAAGCTTCTTGAAGACAAACAACGCTCTGTAGCGACCCCTTTCAGGGCAGCGGAAC
CCCCACCTGGCGACAGGTGCCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCA
AAGGCGGCACAACCCACGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGATCAAATGG
CTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTTGTATG
GGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTGTAGTCGAGGTTAAAAAAA
CGTCTAGGCCCCCGAACACGGGGACGTGGTTTTCTTTGAAAAACACGATAATACCAT
GGGCGCCCCGCGCCAGCGTGCTGAGCGGCGGCGAGCTGGACAAGTGGGAGAAGATCCGCCT
GCGCCCCGCGCGCAAGAGAAGTACAAGCTGAAGCACATCGTGTGGGCCAGCCGCGAGCT
GGAGCGCTTTCGCGGTGAACCCCGGCCTGCTGGAGACCAGCGAGGGCTGCCGCCAGATCCT
GGGCCAGCTGCAGCCCGACCTGCAGACCGGCAGCGAGGAGCTGCGCAGCCTGTACAACAC
CGTGGCCACCTGTACTGCGTGCACCAGCGCATCGACGTCGAAGACACCAAGGAGGGCCCT
GGAGAAGATCGAGGAGGAGCAGAAACAAGTCCAAGAAGAAGGCCCAGCAGGCCCGCCCGC
CGCCGGCACCCGCAACAGCAGCCAGGTGAGCCAGAATACCCCATCGTGCAGAACCTTGCA
GGGCCAGATGGTGCACCAGGCCATCAGCCCCCGCACCTGAACGCTTGGGTGAAGGTGGT
GGAGGAGAAGGCCCTTCAGCCCCGAGGTGATCCCCATGTTAGCGCCCTGAGCGAGGGCGC
CACCCCCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCCACAGGCCCGCATGCA
GATGCTGAAGGAGACCATCAACGAGGAGGCCGCGAGTGGGACCGCTGCACCCCGTGCA
CGCCGGCCCCATCGCCCCGCGCCAGATGCGCGAGCCCCGCGGCAGCGACATCGCCGGCAC
CACCAGCACCTCGAGGAGCAGATCGGCTGGATGACCAACAACCCCCCATCCCCGTGGG
CGAGATCTACAAGCGGTGGATCATCTGGGCCCTGAACAGATCGTGGCATGTACAGCCC
CACCAGCATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCGCGCACTACGTGGACCG
CTTCTACAAGACCTTCGCGCTGAGCAGGCCAGCCAGGACGTGAAGAACTGGATGACCGA
GACCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAGACCATCTGAAGGCTCTCGGCC
CGCGGCCACCTTGAGGAGATGATGACCGCTGCCAGGGCGTGGGCGGCCCGGCCACAA
GGCCCCGCTGCTGGCCGAGCGCATGAGCCAGGTGACGAACCCGCGCACCATCATGATGCA
GCGCGGCAACTTCGCAACAGCGGAAGACCGTCAAGTGCTTCAACTGCGGCAAGGAGGG
CCACACCGCCAGGAAGTCCGCGCCCCCGCAAGAAGGGCTGCTGGCGCTGCGGCCGCA
GGGCCACAGATGAAGGACTGCACCGAGCGCCAGGCCAATCTCTGGGCAAGATCTGGCC
CAGCTACAAGGGCGGCCCGCAACTTCTGCGAGGCGGCCCGAGCCCCAGCCAGGAGCCCAT
CGAGGAGAGCTTCCGCTTCGGCGAGGAGAAGACCCCGCAGCCAGAAGCAGGAGCCCAT
CGACAAGGAGCTGTACCCCTGACCAAGCTGCGCAGCTGTTCCGGCAACGACCCCGCAG
CCAGTAAGAATTCAGACTCGAGCAAGTCTAGA

FIG. 63B

(SEQ ID NO:75)

76 / 131

gp160.modSF162.delV2.gag.modSF2

GAATTTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGTGTGGA
 GCAGTCTTTCGTTTCGCCACGCGCGTGGAGAAGCTGTGGGTGACCGTGTAAGTCTACGCGCGT
 CCGGTGTGGAAGGAGGCCACCCACCCCTGTTCTGCGCCAGCGACGCCAAGGCCATACGAC
 ACCGAGGTGCACACGCTGTGGGCCACCCACGCTGCGTGCCACCGACCCCAACCCCCAG
 GAGATCGTGCTGGAGAAGCTGACCGAGAAGCTTCAACATGTGGAAGAAACAACATGGTGGAG
 CAGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGGTG
 ACCCCCTGTGCGTGACCTGCACTGCACCAACCTGAAGAAGCCACCAACCAAGAGC
 AGCAACTGGAAGGAGATGGACCGCGCGAGATCAAGAAGCTGCAGCTTCAAGGTGGGCGCC
 GGCAGCTGATCAACTGCAACACCAGCGTGATACCCAGGCCCTGCCCAAGGTGAGCTTC
 GAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTCGCCATCCTGAAGTGCAACGAG
 AAGAAGTTCAACGGCAGCGGCCCTGCACCAACGCTGAGCACCCTGACGTGCAACCCAGGC
 ATCCGCCCCGTGGTGAGCACCCAGCTGTGCTGAACGGCAGCCTGGCCGAGGAGGGCGGTG
 GTGATCCGACGCGAGAACTTCAACGACAACGCCAAGACCATCATCGTGACGTGAAGGAG
 AGCGTGGAGATCAACTGCACCCGCCCAACAACAACCCGCAAGAGCATCAACATCGCG
 CCGCGCCGCGCTTCTACGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTGC
 AACATCAGCGCGGAGAAGTGAACAACAACCTGAAGCAGATCGTGACCAAGCTGCAGGCC
 CAGTTTCGGCAACAAGAACCATCGTGTTCAGCAGAGCAGCGCGCGGACCCCGAGATCGTG
 ATGCACAGCTTCAACTGCGCGCGCGAGTTCTTCTACTGCAACAGCACCCAGCTGTTCAAC
 AGCACTGGAACAACACCATCGGCCCAACAACAACCGCACCATCACTGCCCTGCCCTGC
 CGCATCAAGCAGATCATCAACCGCTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCC
 ATCCGCGGCCAGATCCGCTGCAGCAGCAACATCAACCGCCTGCTGCTGACCCGCGACGCG
 GGCAGGAGATCAGCAACACCACCGAGATCTTCGCCCGCGCGCGCGGACATGCGCGAC
 AACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCTGGGCGTGGCC
 CCCACCAAGGCCAAGCGCGCGTGGTGCAGCGCAGAGAAGCGCGCGCTGACCTGGGCGCC
 ATGTTCTGGGCTTCTGGGCGCGCGCGCGCAGCACCATGGGCGCCGCGAGCTGACCCCTG
 ACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTGACGACGAGACAACCTGCTGGCG
 GCCATCGAGGCCCAGCAGCACTGTGTCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAG
 GCCCGCTGCTGGCCGTGGAGCGCTACTGAAGGACAGCAGCTGCTGGGCATCTGGGGC
 TGCAGCGGCAAGCTGATCTGCACACCGCGCTGCCCTGGAACGCCAGCTGGAAGCAACAAG
 AGCTTGGACAGATCTGGAACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGACAAC
 TACCAACAACCTGATCTTACACCTGATCGAGGAGAGCAGAAACAGCAGGAGAAGAACAG
 CAGGAGCTGCTGGAGCTGGACAAGTGGGCGAGCCTGTGAACTGGTTGACATCAGCAAG
 TGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGCGGCTGGTGGGCTGCGCATC
 GTGTTCAACCGTGTGAGCATCGTGAACCGCGTGCGCCAGGGCTACAGCCCCCTGAGCTTC
 CAGACCCGCTTCCCGCCCCCGCGGCCCGACCGCCCCGAGGGCATCGAGGAGGAGGGC
 GGGGAGCGCGACCGCGACCGCAGCAGCCCCCTGGTGACCGGCTGCTGGCCCTGATCTGG
 GACGACCTGGCAGCGCTGTGCCTGTTTCAGCTACCAACCGCTGCGCGAAGCTGATCCTGATC
 GCCGCGCGCATCTGTGAGCTGTGGGCGCGCGCGCTGGGAGGCGCTGAAGTACTGGGGC
 AACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGCGCCGTGAGCTGTTGCAGCGC
 ATCGCCATCGCCGTGGCGGAGGGCAACGACCGCATCATCGAGTGGCCGAGCGCATCGCG
 CGCGCTTCTGACATCCCCCGCGCATCCGCCAGGGCTTCGAGCGCGCCCTGCTGTAA
 CTCGAGCAAGTCTAGAGAAATCCGCCCCCCCCCCCCCCCTCTCCCTCCCCCCCCC
 TAAAGTTACTGAGCGCTTGAATAAGCCGCTGTGCGTTTGTCTATATGTTATT
 TTCCACCATATTGCGCTCTTTGGCAATGTGAGGGCGCGGAAACCTGGGCCCTGTCTTCTT

FIG. 64A

(SEQ ID NO:76)

77 / 131

GACGAGCATTCTTAGGGGTCTTCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGT
CGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGACAAACAAGTCTGTAGCGACCCCT
TTGCAGGCAGCGGAACCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGT
ATAAGATACACCTGCAAGGCGGCACAACCCAGTGCCACGTTGTGAGTTGGATAGTTGT
GGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCGAA
GGTACCCCATTGTATGGGATCTGATCTGGGCTCGGTGCACATGCTTTACATGTGTGTTA
GTCGAGGTTAAAAAACGCTCTAGGCCCCCGAACCCAGGACCGGGACGTGGTTTTCTTTGAAA
AACACGATAATACCATGGGCGCCCGCGCCAGCGTGCTGAGCGGCGGCGAGCTGGACAAGT
GGGAGAAGATCCGCTGCGCCCCGCGGCGCAAGAAGTACAAGCTGAAGCACATCGTGT
GGGCCAGCCGCGAGCTGGAGCGCTTCGCGTGAACCCCGGCTGTGAGACACGCGAGG
GCTGCCGCCAGATCCTGGGCCAGCTGCAGCCAGCCTGCAGACCGGCAGCGAGGAGCTGC
GCAGCCTGTACAACACGTGGCCACCCTGTACTGCGTGACACGCGCATCGACGTCAAGG
ACACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAAGTCCAAGAAGAAGGCC
AGCAGGCCGCGCGCCGCGCGGCCACCGGCAACAGCAGCCAGGTGAGCCAGAATACCCCA
TCGTGCAGAACTGCAGGGCCAGATGGTGCACCAAGGCCATCAGCCCCCGCACCCCTGAACG
CCTGGGTGAAGGTGGTGGAGAGAAGGCCCTCAGCCCCGAGGTGATCCCATGTTTCAGCG
CCCTGAGCGAGGGCGCCACCCCCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCC
ACCAGGCCCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCGCGCGAGTGGGACC
GCGTGACCCCGTGACGCGCGGCCCATCGCCCCGCGCCAGATGCGCGAGCCCCGCGGCA
CGACATCGCGCGCACCAACAGCACCTGCAGGAGCAGATCGGCTGGATGACCAACAACC
CCCCATCCCCGTGGGCGAGATCTACAAGCGGTGGATCATCTGGGCTGAAACAAGATCG
TGCGGATGTACAGCCCCACAGCATCCTGGACATCCGCGAGGGCCCCAAGGAGCCCTCC
CGCAGTACGTGGACCGCTTCTACAAGACCTGCGCGCTGAGCAGGCCAGCCAGGACGTGA
AGAACTGGATGACCGAGACCTTGTGGTGAGAACGCCAACCCGACTGCAAGACCATCC
TGAAGGCTCTCGGCCCGCGGCCACCTGGAGGAGATGATGACCGCTGCCAGGGCGTGC
GCGGCCCGGCCACAAGGCCGCGTGTGGCCGAGGCGATGAGCCAGGTGACGAACCCGG
CGACCATCATGATGCAGCGCGGCAACTTCGCAACACAGCGGAAGACCTCAAGTGCTTCA
ACTGCGGCAAGGAGGGCCACACGCCAGGAACCTGCCGCGCCCCCGCAAGAAGGGCTGCT
GGGCAAGATCTGGCCAGCTACAAGGCGCGCCCCCGCAACTTCTGCAGAGCCGCCCCG
AGCCACCGCCCCCGGAGGAGGCTTCCGCTTGGCGAGGAGAGACCAACCCCGAGCG
AGAAGCAGGAGCCCATCGACRAGGAGCTGTACCCCTGACCGAGCTGCGCAGCCCTTTCG
GCAACGACCCAGCAGCCAGTAAGAATTACGACTCGAGCAAGTCTAGA

FIG. 64B

(SEQ ID NO:76)



FIG. 65C



FIG. 65B

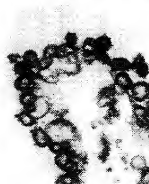


FIG. 65A

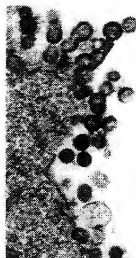


FIG. 65F



FIG. 65E



FIG. 65D

gp160.modSF162	1	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	50
gp160.modSF162.delV2	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp160.modSF162.delV1V2	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp140.modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp140.mut7.modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp140.mut8.modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp120.modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
Consensus		GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	100
gp160.modSF162	51	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCGCGTGGAGAAGCTGTGGG	
gp160.modSF162.delV2	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCGCGTGGAGAAGCTGTGGG	
gp160.modSF162.delV1V2	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCGCGTGGAGAAGCTGTGGG	
gp140.modSF162	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCGCGTGGAGAAGCTGTGGG	
gp140.mut7.modSF162	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCGCGTGGAGAAGCTGTGGG	
gp140.mut8.modSF162	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCGCGTGGAGAAGCTGTGGG	
gp120.modSF162	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCGCGTGGAGAAGCTGTGGG	
Consensus		GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCGCGTGGAGAAGCTGTGGG	100
gp160.modSF162	101	TGACCGTGTAACGCGGTGCCCGTGTGGAGGAGGCCACCCACCCCTG	
gp160.modSF162.delV2	(101)	TGACCGTGTAACGCGGTGCCCGTGTGGAGGAGGCCACCCACCCCTG	
gp160.modSF162.delV1V2	(101)	TGACCGTGTAACGCGGTGCCCGTGTGGAGGAGGCCACCCACCCCTG	
gp140.modSF162	(101)	TGACCGTGTAACGCGGTGCCCGTGTGGAGGAGGCCACCCACCCCTG	
gp140.mut7.modSF162	(101)	TGACCGTGTAACGCGGTGCCCGTGTGGAGGAGGCCACCCACCCCTG	
gp140.mut8.modSF162	(101)	TGACCGTGTAACGCGGTGCCCGTGTGGAGGAGGCCACCCACCCCTG	
gp120.modSF162	(101)	TGACCGTGTAACGCGGTGCCCGTGTGGAGGAGGCCACCCACCCCTG	
Consensus		TGACCGTGTAACGCGGTGCCCGTGTGGAGGAGGCCACCCACCCCTG	100

FIG. 66A-1

80 / 131

[illegible]

FIG. 66A-2

gp120.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAGAACACACATGTTGGAG	350
Consensus	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAGAACACACATGTTGGAG	350
gp160.modSF162	(301)	CAGATGCACGAGGACATCATCAGCTGTGGACCCAGAGCCTGAAGCCCTG	351
gp160.modSF162.delV2	(301)	CAGATGCACGAGGACATCATCAGCTGTGGACCCAGAGCCTGAAGCCCTG	351
gp160.modSF162.delV1V2	(301)	CAGATGCACGAGGACATCATCAGCTGTGGACCCAGAGCCTGAAGCCCTG	351
gp140.modSF162	(301)	CAGATGCACGAGGACATCATCAGCTGTGGACCCAGAGCCTGAAGCCCTG	351
gp140.modSF162	(301)	CAGATGCACGAGGACATCATCAGCTGTGGACCCAGAGCCTGAAGCCCTG	351
gp140.mut.modSF162	(301)	CAGATGCACGAGGACATCATCAGCTGTGGACCCAGAGCCTGAAGCCCTG	351
gp140.mut7.modSF162	(301)	CAGATGCACGAGGACATCATCAGCTGTGGACCCAGAGCCTGAAGCCCTG	351
gp140.mut8.modSF162	(301)	CAGATGCACGAGGACATCATCAGCTGTGGACCCAGAGCCTGAAGCCCTG	351
gp120.modSF162	(301)	CAGATGCACGAGGACATCATCAGCTGTGGACCCAGAGCCTGAAGCCCTG	351
Consensus	(301)	CAGATGCACGAGGACATCATCAGCTGTGGACCCAGAGCCTGAAGCCCTG	351
gp160.modSF162	(351)	CGTGAAGCTGACCCCTGTGGTGACCTCTGCTGCACCAACCTGAAGA	400
gp160.modSF162.delV2	(351)	CGTGAAGCTGACCCCTGTGGTGACCTCTGCTGCACCAACCTGAAGA	400
gp160.modSF162.delV1V2	(351)	CGTGAAGCTGACCCCTGTGGTGACCTCTGCTGCACCAACCTGAAGA	400
gp140.modSF162	(351)	CGTGAAGCTGACCCCTGTGGTGACCTCTGCTGCACCAACCTGAAGA	400
gp140.modSF162	(351)	CGTGAAGCTGACCCCTGTGGTGACCTCTGCTGCACCAACCTGAAGA	400
gp140.mut.modSF162	(351)	CGTGAAGCTGACCCCTGTGGTGACCTCTGCTGCACCAACCTGAAGA	400
gp140.mut7.modSF162	(351)	CGTGAAGCTGACCCCTGTGGTGACCTCTGCTGCACCAACCTGAAGA	400
gp140.mut8.modSF162	(351)	CGTGAAGCTGACCCCTGTGGTGACCTCTGCTGCACCAACCTGAAGA	400
gp120.modSF162	(351)	CGTGAAGCTGACCCCTGTGGTGACCTCTGCTGCACCAACCTGAAGA	400
Consensus	(351)	CGTGAAGCTGACCCCTGTGGTGACCTCTGCTGCACCAACCTGAAGA	400
gp160.modSF162	(401)	ACGCCACCAACCAAGACAGCACTGAGAGGAGATGGACCCGGCGAG	450
gp160.modSF162.delV2	(401)	ACGCCACCAACCAAGACAGCACTGAGAGGAGATGGACCCGGCGAG	450
gp160.modSF162.delV1V2	(375)	ACGCCACCAACCAAGACAGCACTGAGAGGAGATGGACCCGGCGAG	450
gp140.modSF162	(401)	ACGCCACCAACCAAGACAGCACTGAGAGGAGATGGACCCGGCGAG	450
gp140.mut.modSF162	(401)	ACGCCACCAACCAAGACAGCACTGAGAGGAGATGGACCCGGCGAG	450
gp140.mut7.modSF162	(401)	ACGCCACCAACCAAGACAGCACTGAGAGGAGATGGACCCGGCGAG	450
gp140.mut8.modSF162	(401)	ACGCCACCAACCAAGACAGCACTGAGAGGAGATGGACCCGGCGAG	450
gp120.modSF162	(401)	ACGCCACCAACCAAGACAGCACTGAGAGGAGATGGACCCGGCGAG	450
Consensus	(401)	ACGCCACCAACCAAGACAGCACTGAGAGGAGATGGACCCGGCGAG	450

FIG. 66A-3

82 / 131

451	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	500
(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
(376)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
	Consensus	501
gp160.modSF162	gp160.modSF162	
gp160.modSF162.delV2	gp160.modSF162	
gp160.modSF162.delV1V2	gp160.modSF162	
gp140.modSF162	gp140.modSF162	
gp140.mut7.modSF162	gp140.mut7.modSF162	
gp140.mut8.modSF162	gp140.mut8.modSF162	
gp120.modSF162	gp120.modSF162	
	Consensus	550
gp160.modSF162	gp160.modSF162	
gp160.modSF162.delV2	gp160.modSF162	
gp160.modSF162.delV1V2	gp160.modSF162	
gp140.modSF162	gp140.modSF162	
gp140.mut7.modSF162	gp140.mut7.modSF162	
gp140.mut8.modSF162	gp140.mut8.modSF162	
gp120.modSF162	gp120.modSF162	
	Consensus	600
gp160.modSF162	gp160.modSF162	
gp160.modSF162.delV2	gp160.modSF162	
gp160.modSF162.delV1V2	gp160.modSF162	
gp140.modSF162	gp140.modSF162	
gp140.mut7.modSF162	gp140.mut7.modSF162	
gp140.mut8.modSF162	gp140.mut8.modSF162	
gp120.modSF162	gp120.modSF162	
	Consensus	650

FIG. 66A-4

gp140.mut7.modSF162	(551)	ACAACACAGCTCAAGCTGATCAACTGCACACACCGGTGATCCACCG
gp140.mut8.modSF162	(551)	ACAACACAGCTCAAGCTGATCAACTGCACACACCGGTGATCCACCG
gp120.modSF162	(551)	ACAACACAGCTCAAGCTGATCAACTGCACACACCGGTGATCCACCG
Consensus	(551)	ACAACACAGCTCAAGCTGATCAACTGCACACACCGGTGATCCACCG
gp160.modSF162	(601)	GCCTGCCCAAGGTAGCTTCGAGGCCATCCCATCCACTACTGCGCCCG
gp160.modSF162.delV2	(601)	GCCTGCCCAAGGTAGCTTCGAGGCCATCCCATCCACTACTGCGCCCG
gp140.modSF162	(601)	GCCTGCCCAAGGTAGCTTCGAGGCCATCCCATCCACTACTGCGCCCG
gp140.mut7.modSF162	(601)	GCCTGCCCAAGGTAGCTTCGAGGCCATCCCATCCACTACTGCGCCCG
gp140.mut8.modSF162	(601)	GCCTGCCCAAGGTAGCTTCGAGGCCATCCCATCCACTACTGCGCCCG
gp120.modSF162	(601)	GCCTGCCCAAGGTAGCTTCGAGGCCATCCCATCCACTACTGCGCCCG
Consensus	(601)	GCCTGCCCAAGGTAGCTTCGAGGCCATCCCATCCACTACTGCGCCCG
gp160.modSF162	(651)	CGCGGCTTCGCCATCCTGAAGTGCACGACGACGAGGTTCAACGGCAGCG
gp160.modSF162.delV2	(651)	CGCGGCTTCGCCATCCTGAAGTGCACGACGACGAGGTTCAACGGCAGCG
gp160.modSF162.delV1V2	(651)	CGCGGCTTCGCCATCCTGAAGTGCACGACGACGAGGTTCAACGGCAGCG
gp140.modSF162	(651)	CGCGGCTTCGCCATCCTGAAGTGCACGACGACGAGGTTCAACGGCAGCG
gp140.mut7.modSF162	(651)	CGCGGCTTCGCCATCCTGAAGTGCACGACGACGAGGTTCAACGGCAGCG
gp140.mut8.modSF162	(651)	CGCGGCTTCGCCATCCTGAAGTGCACGACGACGAGGTTCAACGGCAGCG
gp120.modSF162	(651)	CGCGGCTTCGCCATCCTGAAGTGCACGACGACGAGGTTCAACGGCAGCG
Consensus	(651)	CGCGGCTTCGCCATCCTGAAGTGCACGACGACGAGGTTCAACGGCAGCG
gp160.modSF162	(701)	GCCCTGCACCAAGTGCAGCCGTCAGTGCACCCACGCGCATCCGCGCC
gp160.modSF162.delV2	(701)	GCCCTGCACCAAGTGCAGCCGTCAGTGCACCCACGCGCATCCGCGCC
gp160.modSF162.delV1V2	(701)	GCCCTGCACCAAGTGCAGCCGTCAGTGCACCCACGCGCATCCGCGCC
gp140.modSF162	(701)	GCCCTGCACCAAGTGCAGCCGTCAGTGCACCCACGCGCATCCGCGCC
gp140.mut7.modSF162	(701)	GCCCTGCACCAAGTGCAGCCGTCAGTGCACCCACGCGCATCCGCGCC
gp140.mut8.modSF162	(701)	GCCCTGCACCAAGTGCAGCCGTCAGTGCACCCACGCGCATCCGCGCC
gp120.modSF162	(701)	GCCCTGCACCAAGTGCAGCCGTCAGTGCACCCACGCGCATCCGCGCC
Consensus	(701)	GCCCTGCACCAAGTGCAGCCGTCAGTGCACCCACGCGCATCCGCGCC

FIG. 66A-5

84 / 131

	751	GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGCGGT	800
gp160.modSF162	(751)	GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGCGGT	
gp160.modSF162.delV2	(670)	GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGCGGT	
gp160.modSF162.delV1V2	(562)	GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGCGGT	
gp140.modSF162	(751)	GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGCGGT	
gp140.modSF162	(751)	GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGCGGT	
gp140.mut7.modSF162	(751)	GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGCGGT	
gp140.mut8.modSF162	(751)	GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGCGGT	
gp120.modSF162	(751)	GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGCGGT	
Consensus		GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGCGGT	850
gp160.modSF162	(801)	GTGTATCCGACGCGAGAACTTCACCGACACGCCAAGCACCATCATCTGTGC	
gp160.modSF162.delV2	(801)	GTGTATCCGACGCGAGAACTTCACCGACACGCCAAGCACCATCATCTGTGC	
gp160.modSF162.delV1V2	(612)	GTGTATCCGACGCGAGAACTTCACCGACACGCCAAGCACCATCATCTGTGC	
gp140.modSF162	(801)	GTGTATCCGACGCGAGAACTTCACCGACACGCCAAGCACCATCATCTGTGC	
gp140.modSF162	(801)	GTGTATCCGACGCGAGAACTTCACCGACACGCCAAGCACCATCATCTGTGC	
gp140.mut7.modSF162	(801)	GTGTATCCGACGCGAGAACTTCACCGACACGCCAAGCACCATCATCTGTGC	
gp140.mut8.modSF162	(801)	GTGTATCCGACGCGAGAACTTCACCGACACGCCAAGCACCATCATCTGTGC	
gp120.modSF162	(801)	GTGTATCCGACGCGAGAACTTCACCGACACGCCAAGCACCATCATCTGTGC	
Consensus		GTGTATCCGACGCGAGAACTTCACCGACACGCCAAGCACCATCATCTGTGC	900
gp160.modSF162	(851)	AGCTGAAGGAGACGCTGGAGATCAACTGCACCCGCCGCCCAACAACACACC	
gp160.modSF162.delV2	(851)	AGCTGAAGGAGACGCTGGAGATCAACTGCACCCGCCGCCCAACAACACACC	
gp160.modSF162.delV1V2	(770)	AGCTGAAGGAGACGCTGGAGATCAACTGCACCCGCCGCCCAACAACACACC	
	(662)	AGCTGAAGGAGACGCTGGAGATCAACTGCACCCGCCGCCCAACAACACACC	

FIG. 66A-6

gp140.modSF162	(851)	AGCTGAAGAGAGCGTGAGATCAACTGACCGCGCGCCCAACAACACC	950
gp140.mut.modSF162	(851)	AGCTGAAGAGAGCGTGAGATCAACTGACCGCGCGCCCAACAACACC	
gp140.mut7.modSF162	(851)	AGCTGAAGAGAGCGTGAGATCAACTGACCGCGCGCCCAACAACACC	
gp140.mut8.modSF162	(851)	AGCTGAAGAGAGCGTGAGATCAACTGACCGCGCGCCCAACAACACC	
gp120.modSF162	(851)	AGCTGAAGAGAGCGTGAGATCAACTGACCGCGCGCCCAACAACACC	
Consensus		901	
gp160.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCGCGCGCGCTTCTAGCGCCACGGCGA	
gp160.modSF162.delV2	(820)	CGCAAGAGCATCACCATCGGCCCGCGCGCGCTTCTAGCGCCACGGCGA	
gp160.modSF162.delV1V2	(712)	CGCAAGAGCATCACCATCGGCCCGCGCGCGCTTCTAGCGCCACGGCGA	
gp140.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCGCGCGCGCTTCTAGCGCCACGGCGA	
gp140.mut.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCGCGCGCGCTTCTAGCGCCACGGCGA	
gp140.mut7.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCGCGCGCGCTTCTAGCGCCACGGCGA	
gp140.mut8.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCGCGCGCGCTTCTAGCGCCACGGCGA	
gp120.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCGCGCGCGCTTCTAGCGCCACGGCGA	
Consensus		951	
gp160.modSF162	(951)	CATCATCGCGGACATCCGCCAGGCCCACTGCAACATCAGCGCGGAGAAGT	
gp160.modSF162.delV2	(870)	CATCATCGCGGACATCCGCCAGGCCCACTGCAACATCAGCGCGGAGAAGT	
gp160.modSF162.delV1V2	(762)	CATCATCGCGGACATCCGCCAGGCCCACTGCAACATCAGCGCGGAGAAGT	
gp140.modSF162	(951)	CATCATCGCGGACATCCGCCAGGCCCACTGCAACATCAGCGCGGAGAAGT	
gp140.mut.modSF162	(951)	CATCATCGCGGACATCCGCCAGGCCCACTGCAACATCAGCGCGGAGAAGT	
gp140.mut7.modSF162	(951)	CATCATCGCGGACATCCGCCAGGCCCACTGCAACATCAGCGCGGAGAAGT	
gp140.mut8.modSF162	(951)	CATCATCGCGGACATCCGCCAGGCCCACTGCAACATCAGCGCGGAGAAGT	
gp120.modSF162	(951)	CATCATCGCGGACATCCGCCAGGCCCACTGCAACATCAGCGCGGAGAAGT	
Consensus		951	

FIG. 66A-7

1001	GGACACACCCCTGAAGCAGATCGTGTACCAAGCTGCAGGCCAGTTCGGC	(1001)	1050
gp160.modSF162	GGACACACCCCTGAAGCAGATCGTGTACCAAGCTGCAGGCCAGTTCGGC	(1001)	
gp160.modSF162.delV2	GGACACACCCCTGAAGCAGATCGTGTACCAAGCTGCAGGCCAGTTCGGC	(920)	
gp160.modSF162.delV1V2	GGACACACCCCTGAAGCAGATCGTGTACCAAGCTGCAGGCCAGTTCGGC	(812)	
gp140.modSF162	GGACACACCCCTGAAGCAGATCGTGTACCAAGCTGCAGGCCAGTTCGGC	(1001)	
gp140.modSF162	GGACACACCCCTGAAGCAGATCGTGTACCAAGCTGCAGGCCAGTTCGGC	(1001)	
gp140.mut.modSF162	GGACACACCCCTGAAGCAGATCGTGTACCAAGCTGCAGGCCAGTTCGGC	(1001)	
gp140.mut7.modSF162	GGACACACCCCTGAAGCAGATCGTGTACCAAGCTGCAGGCCAGTTCGGC	(1001)	
gp140.mut8.modSF162	GGACACACCCCTGAAGCAGATCGTGTACCAAGCTGCAGGCCAGTTCGGC	(1001)	
gp120.modSF162	GGACACACCCCTGAAGCAGATCGTGTACCAAGCTGCAGGCCAGTTCGGC	(1001)	
Consensus		1051	
gp160.modSF162	AACAAGACCATCTGTTCAAGCAGACGACGGCGGCGGCCGAGATCGT	(1051)	
gp160.modSF162.delV2	AACAAGACCATCTGTTCAAGCAGACGACGGCGGCGGCCGAGATCGT	(970)	
gp160.modSF162.delV1V2	AACAAGACCATCTGTTCAAGCAGACGACGGCGGCGGCCGAGATCGT	(862)	
gp140.modSF162	AACAAGACCATCTGTTCAAGCAGACGACGGCGGCGGCCGAGATCGT	(1051)	
gp140.mut.modSF162	AACAAGACCATCTGTTCAAGCAGACGACGGCGGCGGCCGAGATCGT	(1051)	
gp140.mut7.modSF162	AACAAGACCATCTGTTCAAGCAGACGACGGCGGCGGCCGAGATCGT	(1051)	
gp140.mut8.modSF162	AACAAGACCATCTGTTCAAGCAGACGACGGCGGCGGCCGAGATCGT	(1051)	
gp120.modSF162	AACAAGACCATCTGTTCAAGCAGACGACGGCGGCGGCCGAGATCGT	(1051)	
Consensus		1101	
gp160.modSF162	GATGCACAGCTTCAACTGCGGCGGAGTTCCTTCTACTGCACACGACCC	(1101)	
gp160.modSF162.delV2	GATGCACAGCTTCAACTGCGGCGGAGTTCCTTCTACTGCACACGACCC	(1020)	
gp160.modSF162.delV1V2	GATGCACAGCTTCAACTGCGGCGGAGTTCCTTCTACTGCACACGACCC	(912)	
gp140.modSF162	GATGCACAGCTTCAACTGCGGCGGAGTTCCTTCTACTGCACACGACCC	(1101)	
gp140.mut.modSF162	GATGCACAGCTTCAACTGCGGCGGAGTTCCTTCTACTGCACACGACCC	(1101)	
gp140.mut7.modSF162	GATGCACAGCTTCAACTGCGGCGGAGTTCCTTCTACTGCACACGACCC	(1101)	
gp140.mut8.modSF162	GATGCACAGCTTCAACTGCGGCGGAGTTCCTTCTACTGCACACGACCC	(1101)	
gp120.modSF162	GATGCACAGCTTCAACTGCGGCGGAGTTCCTTCTACTGCACACGACCC	(1101)	
Consensus		1151	
gp160.modSF162	AGCTGTTCAACAGCAGCTTGAACACACCATCGCGGCCCCCAACACACCAAC	(1151)	

FIG. 66A-8

gp160.modSF162.delV2	(1070)	AGCTGTTCAACAGCAGCACTGGAGCAACACCACTGGCCCAACACCAAC
gp160.modSF162.delV1V2	(962)	AGCTGTTCAACAGCAGCACTGGAGCAACACCACTGGCCCAACACCAAC
gp140.modSF162	(1151)	AGCTGTTCAACAGCAGCACTGGAGCAACACCACTGGCCCAACACCAAC
gp140.mut.modSF162	(1151)	AGCTGTTCAACAGCAGCACTGGAGCAACACCACTGGCCCAACACCAAC
gp140.mut7.modSF162	(1151)	AGCTGTTCAACAGCAGCACTGGAGCAACACCACTGGCCCAACACCAAC
gp140.mut8.modSF162	(1151)	AGCTGTTCAACAGCAGCACTGGAGCAACACCACTGGCCCAACACCAAC
gp120.modSF162	(1151)	AGCTGTTCAACAGCAGCACTGGAGCAACACCACTGGCCCAACACCAAC
Consensus	1201	
gp160.modSF162	(1201)	GGCACCATCACCCCTGCCCTGGCGCATCAAGCAGATCATCAACCGGTGGCA
gp160.modSF162.delV2	(1120)	GGCACCATCACCCCTGCCCTGGCGCATCAAGCAGATCATCAACCGGTGGCA
gp160.modSF162.delV1V2	(1012)	GGCACCATCACCCCTGCCCTGGCGCATCAAGCAGATCATCAACCGGTGGCA
gp140.modSF162	(1201)	GGCACCATCACCCCTGCCCTGGCGCATCAAGCAGATCATCAACCGGTGGCA
gp140.mut.modSF162	(1201)	GGCACCATCACCCCTGCCCTGGCGCATCAAGCAGATCATCAACCGGTGGCA
gp140.mut7.modSF162	(1201)	GGCACCATCACCCCTGCCCTGGCGCATCAAGCAGATCATCAACCGGTGGCA
gp140.mut8.modSF162	(1201)	GGCACCATCACCCCTGCCCTGGCGCATCAAGCAGATCATCAACCGGTGGCA
gp120.modSF162	(1201)	GGCACCATCACCCCTGCCCTGGCGCATCAAGCAGATCATCAACCGGTGGCA
Consensus	1251	
gp160.modSF162	(1251)	GGAGGTGGGCAAGGCCATGTAGCGCCCGCCCATCCGGGCCAGATCCGCT
gp160.modSF162.delV2	(1170)	GGAGGTGGGCAAGGCCATGTAGCGCCCGCCCATCCGGGCCAGATCCGCT
gp160.modSF162.delV1V2	(1062)	GGAGGTGGGCAAGGCCATGTAGCGCCCGCCCATCCGGGCCAGATCCGCT
gp140.modSF162	(1251)	GGAGGTGGGCAAGGCCATGTAGCGCCCGCCCATCCGGGCCAGATCCGCT
gp140.mut.modSF162	(1251)	GGAGGTGGGCAAGGCCATGTAGCGCCCGCCCATCCGGGCCAGATCCGCT
gp140.mut7.modSF162	(1251)	GGAGGTGGGCAAGGCCATGTAGCGCCCGCCCATCCGGGCCAGATCCGCT
gp140.mut8.modSF162	(1251)	GGAGGTGGGCAAGGCCATGTAGCGCCCGCCCATCCGGGCCAGATCCGCT
gp120.modSF162	(1251)	GGAGGTGGGCAAGGCCATGTAGCGCCCGCCCATCCGGGCCAGATCCGCT
Consensus	1251	

FIG. 66A-9

gp160.modSF162	(1301)	GCAGCAGCAACATCACCGGCGCTGCTGCTGACCCCGCAGCGCGCGCAAGGAG	1350
gp160.modSF162.delV2	(1220)	GCAGCAGCAACATCACCGGCGCTGCTGCTGACCCCGCAGCGCGCGCAAGGAG	
gp160.modSF162.delV1V2	(1112)	GCAGCAGCAACATCACCGGCGCTGCTGCTGACCCCGCAGCGCGCGCAAGGAG	
gp140.modSF162	(1301)	GCAGCAGCAACATCACCGGCGCTGCTGCTGACCCCGCAGCGCGCGCAAGGAG	
gp140.mut.modSF162	(1301)	GCAGCAGCAACATCACCGGCGCTGCTGCTGACCCCGCAGCGCGCGCAAGGAG	
gp140.mut7.modSF162	(1301)	GCAGCAGCAACATCACCGGCGCTGCTGCTGACCCCGCAGCGCGCGCAAGGAG	
gp140.mut8.modSF162	(1301)	GCAGCAGCAACATCACCGGCGCTGCTGCTGACCCCGCAGCGCGCGCAAGGAG	
gp120.modSF162	(1301)	GCAGCAGCAACATCACCGGCGCTGCTGCTGACCCCGCAGCGCGCGCAAGGAG	
Consensus	(1301)	GCAGCAGCAACATCACCGGCGCTGCTGCTGACCCCGCAGCGCGCGCAAGGAG	1400
gp160.modSF162	(1351)	ATCAGCAACACACCCGAGATCTTCGCGCCCGCGCGCGCGACATGCGCGA	
gp160.modSF162.delV2	(1270)	ATCAGCAACACACCCGAGATCTTCGCGCCCGCGCGCGCGACATGCGCGA	
gp160.modSF162.delV1V2	(1162)	ATCAGCAACACACCCGAGATCTTCGCGCCCGCGCGCGCGACATGCGCGA	
gp140.modSF162	(1351)	ATCAGCAACACACCCGAGATCTTCGCGCCCGCGCGCGCGACATGCGCGA	
gp140.mut.modSF162	(1351)	ATCAGCAACACACCCGAGATCTTCGCGCCCGCGCGCGCGACATGCGCGA	
gp140.mut7.modSF162	(1351)	ATCAGCAACACACCCGAGATCTTCGCGCCCGCGCGCGCGACATGCGCGA	
gp140.mut8.modSF162	(1351)	ATCAGCAACACACCCGAGATCTTCGCGCCCGCGCGCGCGACATGCGCGA	
gp120.modSF162	(1351)	ATCAGCAACACACCCGAGATCTTCGCGCCCGCGCGCGCGACATGCGCGA	
Consensus	(1351)	ATCAGCAACACACCCGAGATCTTCGCGCCCGCGCGCGCGACATGCGCGA	1450
gp160.modSF162	(1401)	CAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCC	
gp160.modSF162.delV2	(1320)	CAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCC	
gp160.modSF162.delV1V2	(1212)	CAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCC	
gp140.modSF162	(1401)	CAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCC	
gp140.mut.modSF162	(1401)	CAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCC	
gp140.mut7.modSF162	(1401)	CAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCC	
gp140.mut8.modSF162	(1401)	CAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCC	
gp120.modSF162	(1401)	CAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCC	
Consensus	(1401)	CAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCC	

FIG. 66A-10

gp160.modSF162	(1601)	TGCTGAGCGGCATCGTCAGCAGCAGAACAACTGCTGGCGCCATCGAG	1601
gp160.modSF162.delV2	(1520)	TGCTGAGCGGCATCGTCAGCAGCAGAACAACTGCTGGCGCCATCGAG	1650
gp160.modSF162.delV1V2	(1412)	TGCTGAGCGGCATCGTCAGCAGCAGAACAACTGCTGGCGCCATCGAG	1700
gp140.modSF162	(1601)	TGCTGAGCGGCATCGTCAGCAGCAGAACAACTGCTGGCGCCATCGAG	1651
gp140.modSF162	(1601)	TGCTGAGCGGCATCGTCAGCAGCAGAACAACTGCTGGCGCCATCGAG	1701
gp140.mut7.modSF162	(1601)	TGCTGAGCGGCATCGTCAGCAGCAGAACAACTGCTGGCGCCATCGAG	1651
gp140.mut8.modSF162	(1601)	TGCTGAGCGGCATCGTCAGCAGCAGAACAACTGCTGGCGCCATCGAG	1701
gp120.modSF162	(1513)	TGCTGAGCGGCATCGTCAGCAGCAGAACAACTGCTGGCGCCATCGAG	1651
Consensus	(1601)	TGCTGAGCGGCATCGTCAGCAGCAGAACAACTGCTGGCGCCATCGAG	1701
gp160.modSF162	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGCATCAAGCAGCTGCA	1651
gp160.modSF162.delV2	(1570)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGCATCAAGCAGCTGCA	1700
gp160.modSF162.delV1V2	(1462)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGCATCAAGCAGCTGCA	1651
gp140.modSF162	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGCATCAAGCAGCTGCA	1701
gp140.mut.modSF162	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGCATCAAGCAGCTGCA	1651
gp140.mut7.modSF162	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGCATCAAGCAGCTGCA	1701
gp140.mut8.modSF162	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGCATCAAGCAGCTGCA	1651
gp120.modSF162	(1513)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGCATCAAGCAGCTGCA	1701
Consensus	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGCATCAAGCAGCTGCA	1701
gp160.modSF162	(1701)	GCCCGCGTGTGGCGTGGAGCGCTACCTGAAGACCAGCAGCTGCTGG	1651
gp160.modSF162.delV2	(1620)	GCCCGCGTGTGGCGTGGAGCGCTACCTGAAGACCAGCAGCTGCTGG	1700
gp160.modSF162.delV1V2	(1512)	GCCCGCGTGTGGCGTGGAGCGCTACCTGAAGACCAGCAGCTGCTGG	1651
gp140.modSF162	(1701)	GCCCGCGTGTGGCGTGGAGCGCTACCTGAAGACCAGCAGCTGCTGG	1701
gp140.mut.modSF162	(1701)	GCCCGCGTGTGGCGTGGAGCGCTACCTGAAGACCAGCAGCTGCTGG	1651
gp140.mut7.modSF162	(1701)	GCCCGCGTGTGGCGTGGAGCGCTACCTGAAGACCAGCAGCTGCTGG	1701
gp140.mut8.modSF162	(1701)	GCCCGCGTGTGGCGTGGAGCGCTACCTGAAGACCAGCAGCTGCTGG	1651

FIG. 66A-12

gp120.modSF162	(1513)	-----	GGCCCGCGTGTGGCCGTGGAGCGCTACCTGAAGGACCAAGCAGCTGCTGG	1800
Consensus	(1701)	1751	-----	
gp160.modSF162	(1751)	GCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCAACCCGCGTCCCTGG		
gp160.modSF162.delV2	(1670)	GCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCAACCCGCGTCCCTGG		
gp160.modSF162.delV1V2	(1562)	GCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCAACCCGCGTCCCTGG		
gp140.modSF162	(1751)	GCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCAACCCGCGTCCCTGG		
gp140.modSF162	(1751)	GCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCAACCCGCGTCCCTGG		
gp140.mut7.modSF162	(1751)	GCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCAACCCGCGTCCCTGG		
gp140.mut8.modSF162	(1751)	GCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCAACCCGCGTCCCTGG		
gp120.modSF162	(1513)	-----	GCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCAACCCGCGTCCCTGG	
Consensus	(1751)	1801	-----	
gp160.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGATCTGGAACAACATGAC		
gp160.modSF162.delV2	(1720)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGATCTGGAACAACATGAC		
gp160.modSF162.delV1V2	(1612)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGATCTGGAACAACATGAC		
gp140.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGATCTGGAACAACATGAC		
gp140.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGATCTGGAACAACATGAC		
gp140.mut7.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGATCTGGAACAACATGAC		
gp140.mut8.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGATCTGGAACAACATGAC		
gp120.modSF162	(1513)	-----	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGATCTGGAACAACATGAC	
Consensus	(1801)	1851	-----	
gp160.modSF162	(1851)	CTGGATGGAGTGGGAGCGGAGATCGACAACCTACACCAACCTGATCTACA		
gp160.modSF162.delV2	(1770)	CTGGATGGAGTGGGAGCGGAGATCGACAACCTACACCAACCTGATCTACA		
gp160.modSF162.delV1V2	(1662)	CTGGATGGAGTGGGAGCGGAGATCGACAACCTACACCAACCTGATCTACA		
gp140.modSF162	(1851)	CTGGATGGAGTGGGAGCGGAGATCGACAACCTACACCAACCTGATCTACA		
gp140.mut7.modSF162	(1851)	CTGGATGGAGTGGGAGCGGAGATCGACAACCTACACCAACCTGATCTACA		
gp140.mut8.modSF162	(1851)	CTGGATGGAGTGGGAGCGGAGATCGACAACCTACACCAACCTGATCTACA		
gp120.modSF162	(1513)	-----	CTGGATGGAGTGGGAGCGGAGATCGACAACCTACACCAACCTGATCTACA	
Consensus	(1851)	1900	-----	

FIG. 66A-13

1901	gp160.modSF162	(1901)	CCCTGATCGAGGAGGCGAGAACCCAGCAGGAGAGAACCGAGCAGGAGCTG	1950
	gp160.modSF162.delV2	(1820)	CCCTGATCGAGGAGGCGAGAACCCAGCAGGAGAGAGAACCGAGCAGGAGCTG	
	gp160.modSF162.delV1V2	(1712)	CCCTGATCGAGGAGGCGAGAACCCAGCAGGAGAGAGAACCGAGCAGGAGCTG	
	gp140.modSF162	(1901)	CCCTGATCGAGGAGGCGAGAACCCAGCAGGAGAGAGAACCGAGCAGGAGCTG	
	gp140.modSF162	(1901)	CCCTGATCGAGGAGGCGAGAACCCAGCAGGAGAGAGAACCGAGCAGGAGCTG	
	gp140.mut7.modSF162	(1901)	CCCTGATCGAGGAGGCGAGAACCCAGCAGGAGAGAGAACCGAGCAGGAGCTG	
	gp140.mut8.modSF162	(1901)	CCCTGATCGAGGAGGCGAGAACCCAGCAGGAGAGAGAACCGAGCAGGAGCTG	
	gp120.modSF162	(1513)	-----	
	Consensus	(1901)	CCCTGATCGAGGAGGCGAGAACCCAGCAGGAGAGAGAACCGAGCAGGAGCTG	
	gp160.modSF162	(1951)	CTGGAGCTGGACAAAGTGGGCCAGGCTGTGGAACTGGTTGCACATCAGCAA	2050
	gp160.modSF162.delV2	(1870)	CTGGAGCTGGACAAAGTGGGCCAGGCTGTGGAACTGGTTGCACATCAGCAA	
	gp160.modSF162.delV1V2	(1762)	CTGGAGCTGGACAAAGTGGGCCAGGCTGTGGAACTGGTTGCACATCAGCAA	
	gp140.modSF162	(1951)	CTGGAGCTGGACAAAGTGGGCCAGGCTGTGGAACTGGTTGCACATCAGCAA	
	gp140.modSF162	(1951)	CTGGAGCTGGACAAAGTGGGCCAGGCTGTGGAACTGGTTGCACATCAGCAA	
	gp140.mut7.modSF162	(1951)	CTGGAGCTGGACAAAGTGGGCCAGGCTGTGGAACTGGTTGCACATCAGCAA	
	gp140.mut8.modSF162	(1951)	CTGGAGCTGGACAAAGTGGGCCAGGCTGTGGAACTGGTTGCACATCAGCAA	
	gp120.modSF162	(1513)	-----	
	Consensus	(1951)	CTGGAGCTGGACAAAGTGGGCCAGGCTGTGGAACTGGTTGCACATCAGCAA	
2001	gp160.modSF162	(2001)	GTGGCTGTGGTACATCAAGATCTTCATCATGTCGTGGCGGCTTGGTGG	2050
	gp160.modSF162.delV2	(1920)	GTGGCTGTGGTACATCAAGATCTTCATCATGTCGTGGCGGCTTGGTGG	
	gp160.modSF162.delV1V2	(1812)	GTGGCTGTGGTACATCAAGATCTTCATCATGTCGTGGCGGCTTGGTGG	
	gp140.modSF162	(2001)	GTGGCTGTGGTACATCTAACTCGAG-----	
	gp140.mut.modSF162	(2001)	GTGGCTGTGGTACATCTAACTCGAG-----	

FIG. 66A-14

gp140 .mut7 .modSF162	(2001)	GTGGCTGTGTCATCTTAATCGAG	2100
gp140 .modSF162	(2001)	GTGGCTGTGTCATCTTAATCGAG	
gp120 .modSF162	(1513)	-----	
Consensus	(2001)	GTGGCTGTGTCATCTTAATCGAG	
gp160 .modSF162	(2051)	2051	
gp160 .modSF162 .delV2	(1970)	GCCTGCGCATCGTGTTACCCGCTGCTGAGCATCGTGAACCGCGTGGCGCAG	2100
gp160 .modSF162 .delV1V2	(1862)	GCCTGCGCATCGTGTTACCCGCTGCTGAGCATCGTGAACCGCGTGGCGCAG	
gp140 .modSF162	(2026)	GCCTGCGCATCGTGTTACCCGCTGCTGAGCATCGTGAACCGCGTGGCGCAG	
gp140 .mut .modSF162	(2026)	-----	
gp140 .mut7 .modSF162	(2026)	-----	
gp140 .mut8 .modSF162	(2026)	-----	
gp120 .modSF162	(1513)	-----	
Consensus	(2051)	-----	
gp160 .modSF162	(2101)	2101	
gp160 .modSF162 .delV2	(2020)	GGCTACAGCCCCCTGAGCTTCAGACCCGCTTCCCGCCGCCCGCGGGGCC	2150
gp160 .modSF162 .delV1V2	(1912)	GGCTACAGCCCCCTGAGCTTCAGACCCGCTTCCCGCCGCCCGCGGGGCC	
gp140 .modSF162	(2026)	GGCTACAGCCCCCTGAGCTTCAGACCCGCTTCCCGCCGCCCGCGGGGCC	
gp140 .mut .modSF162	(2026)	-----	
gp140 .mut7 .modSF162	(2026)	-----	
gp140 .mut8 .modSF162	(2026)	-----	
gp120 .modSF162	(1513)	-----	
Consensus	(2101)	-----	
gp160 .modSF162	(2151)	2151	
gp160 .modSF162 .delV2	(2151)	CGACCGCCCGGAGGGCATCGAGGAGGGGGGGCGGAGCGCGACCGGACCC	2200
gp160 .modSF162 .delV1V2	(1962)	CGACCGCCCGGAGGGCATCGAGGAGGGGGGGCGGAGCGCGACCGGACCC	
gp140 .modSF162	(2026)	CGACCGCCCGGAGGGCATCGAGGAGGGGGGGCGGAGCGCGACCGGACCC	
gp140 .mut .modSF162	(2026)	-----	
gp140 .mut7 .modSF162	(2026)	-----	
gp140 .mut8 .modSF162	(2026)	-----	
gp120 .modSF162	(1513)	-----	
Consensus	(2151)	-----	

FIG. 66A-15

gp140.modSF162	(2026)	-----	
gp140.mut.modSF162	(2026)	-----	
gp140.mut7.modSF162	(2026)	-----	
gp140.mut8.modSF162	(2026)	-----	
gp120.modSF162	(1513)	-----	
Consensus	(2301)	-----	
gp160.modSF162	(2351)	-----	2450
gp160.modSF162.delV2	(2351)	AGTACTGGGGCAACCTGCTGCGAGTACTGGATCCAGGAGCTGAGACACGC	
gp160.modSF162.delV1V2	(2370)	AGTACTGGGGCAACCTGCTGCGAGTACTGGATCCAGGAGCTGAGACACGC	
gp140.modSF162	(2162)	AGTACTGGGGCAACCTGCTGCGAGTACTGGATCCAGGAGCTGAGACACGC	
gp140.mut.modSF162	(2026)	-----	
gp140.mut7.modSF162	(2026)	-----	
gp140.mut8.modSF162	(2026)	-----	
gp120.modSF162	(1513)	-----	
Consensus	(2351)	-----	
gp160.modSF162	(2401)	-----	2450
gp160.modSF162.delV2	(2320)	GCCGTGAGCCCTGTTTCGACGCCATCGCCATCGCCGTGGCCGAGGCGCACCGA	
gp160.modSF162.delV1V2	(2212)	GCCGTGAGCCCTGTTTCGACGCCATCGCCATCGCCGTGGCCGAGGCGCACCGA	
gp140.modSF162	(2026)	-----	
gp140.mut.modSF162	(2026)	-----	
gp140.mut7.modSF162	(2026)	-----	
gp140.mut8.modSF162	(2026)	-----	
gp120.modSF162	(1513)	-----	
Consensus	(2401)	-----	

FIG. 66A-17

97 / 131

		Start of tPA	
		1	40
gp160	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT	
gp160 del V1	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGAGGGGCTCTGCT	
gp160 del V2	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGAGGGGCTCTGCT	
gp160 del V1-2	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGAGGGGCTCTGCT	
gp 160 del 128-194	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGAGGGGCTCTGCT	
gp140TM	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGAGGGGCTCTGCT	
gp140	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGAGGGGCTCTGCT	
gp140mut	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGAGGGGCTCTGCT	
gp120	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGAGGGGCTCTGCT	
Consensus	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGAGGGGCTCTGCT	
		41	80
gp160	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG	
gp160 del V1	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG	
gp160 del V2	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG	
gp160 del V1-2	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG	
gp 160 del 128-194	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG	
gp140TM	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG	
gp140	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG	
gp140mut	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG	
gp120	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG	
Consensus	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG	
end of tPA		81	120
		↓	
gp160	(81)	CGCCACCACCGTGCTGTGGGTGACCGGTGTACTACGGCGTG	
gp160 del V1	(81)	CGCCACCACCGTGCTGTGGGTGACCGGTGTACTACGGCGTG	
gp160 del V2	(81)	CGCCACCACCGTGCTGTGGGTGACCGGTGTACTACGGCGTG	
gp160 del V1-2	(81)	CGCCACCACCGTGCTGTGGGTGACCGGTGTACTACGGCGTG	
gp 160 del 128-194	(81)	CGCCACCACCGTGCTGTGGGTGACCGGTGTACTACGGCGTG	
gp140TM	(81)	CGCCACCACCGTGCTGTGGGTGACCGGTGTACTACGGCGTG	
gp140	(81)	CGCCACCACCGTGCTGTGGGTGACCGGTGTACTACGGCGTG	
gp140mut	(81)	CGCCACCACCGTGCTGTGGGTGACCGGTGTACTACGGCGTG	
gp120	(81)	CGCCACCACCGTGCTGTGGGTGACCGGTGTACTACGGCGTG	
Consensus	(81)	CGCCACCACCGTGCTGTGGGTGACCGGTGTACTACGGCGTG	
		121	160
gp 160	(121)	CCCGTGTGGAAGGAGGCCACCCACCACCTGTTCTGCGCCA	
gp160 del V1	(121)	CCCGTGTGGAAGGAGGCCACCCACCACCTGTTCTGCGCCA	
gp160 del V2	(121)	CCCGTGTGGAAGGAGGCCACCCACCACCTGTTCTGCGCCA	
gp160 del V1-2	(121)	CCCGTGTGGAAGGAGGCCACCCACCACCTGTTCTGCGCCA	
gp 160 del 128-194	(121)	CCCGTGTGGAAGGAGGCCACCCACCACCTGTTCTGCGCCA	
gp140TM	(121)	CCCGTGTGGAAGGAGGCCACCCACCACCTGTTCTGCGCCA	
gp140	(121)	CCCGTGTGGAAGGAGGCCACCCACCACCTGTTCTGCGCCA	
gp140mut	(121)	CCCGTGTGGAAGGAGGCCACCCACCACCTGTTCTGCGCCA	
gp120	(121)	CCCGTGTGGAAGGAGGCCACCCACCACCTGTTCTGCGCCA	
Consensus	(121)	CCCGTGTGGAAGGAGGCCACCCACCACCTGTTCTGCGCCA	

FIG. 66B-1

98 / 131

	161	200
gp160	(161) GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp160 del V1	(161) GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp160 del V2	(161) GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp160 del V1-2	(161) GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp 160 del 128-194	(161) GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp140TM	(161) GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp140	(161) GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp140mut	(161) GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp120	(161) GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
Consensus	(161) GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
	201	240
gp160	(201) GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp160 del V1	(201) GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp160 del V2	(201) GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp160 del V1-2	(201) GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp 160 del 128-194	(201) GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp140TM	(201) GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp140	(201) GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp140mut	(201) GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp120	(201) GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
Consensus	(201) GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
	241	280
gp160	(241) GAGGTGAACCTGACCAACGTGACCGAGAAGCTTCAACATGT	
gp160 del V1	(241) GAGGTGAACCTGACCAACGTGACCGAGAAGCTTCAACATGT	
gp160 del V2	(241) GAGGTGAACCTGACCAACGTGACCGAGAAGCTTCAACATGT	
gp160 del V1-2	(241) GAGGTGAACCTGACCAACGTGACCGAGAAGCTTCAACATGT	
gp 160 del 128-194	(241) GAGGTGAACCTGACCAACGTGACCGAGAAGCTTCAACATGT	
gp140TM	(241) GAGGTGAACCTGACCAACGTGACCGAGAAGCTTCAACATGT	
gp140	(241) GAGGTGAACCTGACCAACGTGACCGAGAAGCTTCAACATGT	
gp140mut	(241) GAGGTGAACCTGACCAACGTGACCGAGAAGCTTCAACATGT	
gp120	(241) GAGGTGAACCTGACCAACGTGACCGAGAAGCTTCAACATGT	
Consensus	(241) GAGGTGAACCTGACCAACGTGACCGAGAAGCTTCAACATGT	
	281	320
gp160	(281) GGAAGAACAAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp160 del V1	(281) GGAAGAACAAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp160 del V2	(281) GGAAGAACAAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp160 del V1-2	(281) GGAAGAACAAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp 160 del 128-194	(281) GGAAGAACAAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp140TM	(281) GGAAGAACAAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp140	(281) GGAAGAACAAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp140mut	(281) GGAAGAACAAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp120	(281) GGAAGAACAAACATGGTGGAGCAGATGCATGAGGACATCAT	
Consensus	(281) GGAAGAACAAACATGGTGGAGCAGATGCATGAGGACATCAT	
	321	360
gp160	(321) CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp160 del V1	(321) CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp160 del V2	(321) CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp160 del V1-2	(321) CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp 160 del 128-194	(321) CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp140TM	(321) CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp140	(321) CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp140mut	(321) CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp120	(321) CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
Consensus	(321) CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	

FIG. 66B-2

		99 / 131	
		361	400
gp160	(361)	ACCCCCCTGTGCGTGACCCCTGAAC	GCACCGACAAGCTGA
gp160 del V1	(361)	ACCCCCCTGTGCGTGACCCCTGAAC	GCACCGACAAGCTGG
gp160 del V2	(361)	ACCCCCCTGTGCGTGACCCCTGAAC	GCACCGACAAGCTGA
gp160 del V1-2	(361)	GGC-----	-----
gp 160 del 128-194	(361)	ACCCCCCTGTGCGTGCGGCGCAGG	-----
gp140TM	(361)	ACCCCCCTGTGCGTGACCCCTGAAC	GCACCGACAAGCTGA
gp140	(361)	ACCCCCCTGTGCGTGACCCCTGAAC	GCACCGACAAGCTGA
gp140mut	(361)	ACCCCCCTGTGCGTGACCCCTGAAC	GCACCGACAAGCTGA
gp120	(361)	ACCCCCCTGTGCGTGACCCCTGAAC	GCACCGACAAGCTGA
Consensus	(361)	ACCCCCCTGTGCGTGACCCCTGAAC	GCACCGACAAGCTGA
		401	440
gp160	(401)	CCGGCAGCACCAACGGCACCAACAGC	ACCAGCGGCACCAA
gp160 del V1	(401)	CGGCCGCG-----	-----
gp160 del V2	(401)	CCGGCAGCACCAACGGCACCAACAGC	ACCAGCGGCACCAA
gp160 del V1-2	(364)	-----	-----
gp 160 del 128-194	(385)	-----	-----
gp140TM	(401)	CCGGCAGCACCAACGGCACCAACAGC	ACCAGCGGCACCAA
gp140	(401)	CCGGCAGCACCAACGGCACCAACAGC	ACCAGCGGCACCAA
gp140mut	(401)	CCGGCAGCACCAACGGCACCAACAGC	ACCAGCGGCACCAA
gp120	(401)	CCGGCAGCACCAACGGCACCAACAGC	ACCAGCGGCACCAA
Consensus	(401)	CCGGCAGCACCAACGGCACCAACAGC	ACCAGCGGCACCAA
		441	480
gp160	(441)	CAGCACCAGCGGCACCAACAGCACC	AGCACCACAGCACC
gp160 del V1	(409)	-----	-----
gp160 del V2	(441)	CAGCACCAGCGGCACCAACAGCACC	AGCACCACAGCACC
gp160 del V1-2	(364)	-----	-----
gp 160 del 128-194	(385)	-----	-----
gp140TM	(441)	CAGCACCAGCGGCACCAACAGCACC	AGCACCACAGCACC
gp140	(441)	CAGCACCAGCGGCACCAACAGCACC	AGCACCACAGCACC
gp140mut	(441)	CAGCACCAGCGGCACCAACAGCACC	AGCACCACAGCACC
gp120	(441)	CAGCACCAGCGGCACCAACAGCACC	AGCACCACAGCACC
Consensus	(441)	CAGCACCAGCGGCACCAACAGCACC	AGCACCACAGCACC
		481	520
gp160	(481)	GACAGCTGGGAGAAGATGCCCGAGG	GCGAGATCAAGAACT
gp160 del V1	(409)	-----	GGCGAGATCAAGAACT
gp160 del V2	(481)	GACAGCTGGGAGAAGATGCCCGAGG	GCGAGATCAAGAACT
gp160 del V1-2	(364)	-----	-----
gp 160 del 128-194	(385)	-----	-----
gp140TM	(481)	GACAGCTGGGAGAAGATGCCCGAGG	GCGAGATCAAGAACT
gp140	(481)	GACAGCTGGGAGAAGATGCCCGAGG	GCGAGATCAAGAACT
gp140mut	(481)	GACAGCTGGGAGAAGATGCCCGAGG	GCGAGATCAAGAACT
gp120	(481)	GACAGCTGGGAGAAGATGCCCGAGG	GCGAGATCAAGAACT
Consensus	(481)	GACAGCTGGGAGAAGATGCCCGAGG	GCGAGATCAAGAACT
		521	560
gp160	(521)	GCAGCTTCAACATCACCACACGCGT	GCGCGACAAGGTGCA
gp160 del V1	(521)	GCAGCTTCAACATCACCACACGCGT	GCGCGACAAGGTGCA
gp160 del V2	(521)	GCAGCTTCAACATCGCGCGCGGCG	-----
gp160 del V1-2	(521)	-----	-----
gp 160 del 128-194	(521)	-----	-----
gp140TM	(521)	GCAGCTTCAACATCACCACACGCGT	GCGCGACAAGGTGCA
gp140	(521)	GCAGCTTCAACATCACCACACGCGT	GCGCGACAAGGTGCA
gp140mut	(521)	GCAGCTTCAACATCACCACACGCGT	GCGCGACAAGGTGCA
gp120	(521)	GCAGCTTCAACATCACCACACGCGT	GCGCGACAAGGTGCA
Consensus	(521)	GCAGCTTCAACATCACCACACGCGT	GCGCGACAAGGTGCA

FIG. 66B-3

100 / 131

		561	600
	gp160	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCG
	gp160 del V1	(465)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCG
	gp160 del V2	(544)	-----
	gp160 del V1-2	(364)	-----
	gp 160 del 128-194	(385)	-----
	gp140TM	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCG
	gp140	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCG
	gp140mut	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCG
	gp120	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCG
	Consensus	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCG
		601	640
	gp160	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
	gp160 del V1	(505)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
	gp160 del V2	(544)	-----CGCCTGATCAACTGCA
	gp160 del V1-2	(364)	-----
	gp 160 del 128-194	(385)	-----AACTGCG
	gp140TM	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
	gp140	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
	gp140mut	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
	gp120	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
	Consensus	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
		641	680
	gp160	(641)	ACACCAGCGTGATCACCCAGGCCTGCCCAAGGTGAGCTT
	gp160 del V1	(545)	ACACCAGCGTGATCACCCAGGCCTGCCCAAGGTGAGCTT
	gp160 del V2	(560)	ACACCAGCGTGATCACCCAGGCCTGCCCAAGGTGAGCTT
	gp160 del V1-2	(364)	-----CAGGCCTGCCCAAGGTGAGCTT
	gp 160 del 128-194	(392)	AGACCAGCGTGATCACCCAGGCCTGCCCAAGGTGAGCTT
	gp140TM	(641)	ACACCAGCGTGATCACCCAGGCCTGCCCAAGGTGAGCTT
	gp140	(641)	ACACCAGCGTGATCACCCAGGCCTGCCCAAGGTGAGCTT
	gp140mut	(641)	ACACCAGCGTGATCACCCAGGCCTGCCCAAGGTGAGCTT
	gp120	(641)	ACACCAGCGTGATCACCCAGGCCTGCCCAAGGTGAGCTT
	Consensus	(641)	ACACCAGCGTGATCACCCAGGCCTGCCCAAGGTGAGCTT
		681	720
	gp160	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC
	gp160 del V1	(585)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC
	gp160 del V2	(600)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC
	gp160 del V1-2	(387)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC
	gp 160 del 128-194	(432)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC
	gp140TM	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC
	gp140	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC
	gp140mut	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC
	gp120	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC
	Consensus	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC
		721	760
	gp160	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
	gp160 del V1	(625)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
	gp160 del V2	(640)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
	gp160 del V1-2	(427)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
	gp 160 del 128-194	(472)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
	gp140TM	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
	gp140	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
	gp140mut	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
	gp120	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
	Consensus	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG

FIG. 66B-4

101 / 131

		761	800
gp160	(761)	GCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGG	
gp160 del V1	(665)	GCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGG	
gp160 del V2	(680)	GCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGG	
gp160 del V1-2	(467)	GCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGG	
gp 160 del 128-194	(512)	GCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGG	
gp140TM	(761)	GCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGG	
gp140	(761)	GCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGG	
gp140mut	(761)	GCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGG	
gp120	(761)	GCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGG	
Consensus	(761)	GCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGG	
		801	840
gp160	(801)	CATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp160 del V1	(705)	CATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp160 del V2	(720)	CATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp160 del V1-2	(507)	CATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp 160 del 128-194	(552)	CATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp140TM	(801)	CATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp140	(801)	CATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp140mut	(801)	CATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp120	(801)	CATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
Consensus	(801)	CATCCGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
		841	880
gp160	(841)	AGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAAT	
gp160 del V1	(745)	AGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAAT	
gp160 del V2	(760)	AGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAAT	
gp160 del V1-2	(547)	AGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAAT	
gp 160 del 128-194	(592)	AGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAAT	
gp140TM	(841)	AGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAAT	
gp140	(841)	AGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAAT	
gp140mut	(841)	AGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAAT	
gp120	(841)	AGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAAT	
Consensus	(841)	AGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAAT	
		881	920
gp160	(881)	TCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACGA	
gp160 del V1	(785)	TCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACGA	
gp160 del V2	(800)	TCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACGA	
gp160 del V1-2	(587)	TCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACGA	
gp 160 del 128-194	(632)	TCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACGA	
gp140TM	(881)	TCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACGA	
gp140	(881)	TCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACGA	
gp140mut	(881)	TCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACGA	
gp120	(881)	TCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACGA	
Consensus	(881)	TCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACGA	
		921	960
gp160	(921)	GTCCTGGAGATCAACTGCATCCGCCCAACAACAACACG	
gp160 del V1	(825)	GTCCTGGAGATCAACTGCATCCGCCCAACAACAACACG	
gp160 del V2	(840)	GTCCTGGAGATCAACTGCATCCGCCCAACAACAACACG	
gp160 del V1-2	(627)	GTCCTGGAGATCAACTGCATCCGCCCAACAACAACACG	
gp 160 del 128-194	(672)	GTCCTGGAGATCAACTGCATCCGCCCAACAACAACACG	
gp140TM	(921)	GTCCTGGAGATCAACTGCATCCGCCCAACAACAACACG	
gp140	(921)	GTCCTGGAGATCAACTGCATCCGCCCAACAACAACACG	
gp140mut	(921)	GTCCTGGAGATCAACTGCATCCGCCCAACAACAACACG	
gp120	(921)	GTCCTGGAGATCAACTGCATCCGCCCAACAACAACACG	
Consensus	(921)	GTCCTGGAGATCAACTGCATCCGCCCAACAACAACACG	

FIG. 66B-5

SUBSTITUTE SHEET (RULE 26)

102 / 131

		961	1000
gp160	(961)	CGTAAGAGCATCCACATCGGCCCGGGCGCGCCTTCTACG	
gp160 del V1	(865)	CGTAAGAGCATCCACATCGGCCCGGGCGCGCCTTCTACG	
gp160 del V2	(880)	CGTAAGAGCATCCACATCGGCCCGGGCGCGCCTTCTACG	
gp160 del V1-2	(667)	CGTAAGAGCATCCACATCGGCCCGGGCGCGCCTTCTACG	
gp 160 del 128-194	(712)	CGTAAGAGCATCCACATCGGCCCGGGCGCGCCTTCTACG	
gp140TM	(961)	CGTAAGAGCATCCACATCGGCCCGGGCGCGCCTTCTACG	
gp140	(961)	CGTAAGAGCATCCACATCGGCCCGGGCGCGCCTTCTACG	
gp140mut	(961)	CGTAAGAGCATCCACATCGGCCCGGGCGCGCCTTCTACG	
gp120	(961)	CGTAAGAGCATCCACATCGGCCCGGGCGCGCCTTCTACG	
Consensus	(961)	CGTAAGAGCATCCACATCGGCCCGGGCGCGCCTTCTACG	
		1001	1040
gp160	(1001)	CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp160 del V1	(905)	CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp160 del V2	(920)	CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp160 del V1-2	(707)	CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp 160 del 128-194	(752)	CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp140TM	(1001)	CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp140	(1001)	CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp140mut	(1001)	CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp120	(1001)	CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
Consensus	(1001)	CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
		1041	1080
gp160	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG	
gp160 del V1	(945)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG	
gp160 del V2	(960)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG	
gp160 del V1-2	(747)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG	
gp 160 del 128-194	(792)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG	
gp140TM	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG	
gp140	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG	
gp140mut	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG	
gp120	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG	
Consensus	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG	
		1081	1120
gp160	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGA	
gp160 del V1	(985)	ATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGA	
gp160 del V2	(1000)	ATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGA	
gp160 del V1-2	(787)	ATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGA	
gp 160 del 128-194	(832)	ATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGA	
gp140TM	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGA	
gp140	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGA	
gp140mut	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGA	
gp120	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGA	
Consensus	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGA	
		1121	1160
gp160	(1121)	CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp160 del V1	(1025)	CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp160 del V2	(1040)	CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp160 del V1-2	(827)	CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp 160 del 128-194	(872)	CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp140TM	(1121)	CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp140	(1121)	CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp140mut	(1121)	CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp120	(1121)	CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
Consensus	(1121)	CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	

FIG. 66B-6

SUBSTITUTE SHEET (RULE 26)

103 / 131

		1161	1200
		CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
gp160	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
gp160 del V1	(1065)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
gp160 del V2	(1080)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
gp160 del V1-2	(867)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
gp 160 del 128-194	(912)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
gp140TM	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
gp140	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
gp140mut	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
gp120	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
Consensus	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC	
		1201	1240
		TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
gp160	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
gp160 del V1	(1105)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
gp160 del V2	(1120)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
gp160 del V1-2	(907)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
gp 160 del 128-194	(952)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
gp140TM	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
gp140	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
gp140mut	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
gp120	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
Consensus	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA	
		1241	1280
		CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
gp160	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
gp160 del V1	(1145)	CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
gp160 del V2	(1160)	CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
gp160 del V1-2	(947)	CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
gp 160 del 128-194	(992)	CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
gp140TM	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
gp140	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
gp140mut	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
gp120	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
Consensus	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAAGCACCATCAT	
		1281	1320
		CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp160	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp160 del V1	(1185)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp160 del V2	(1200)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp160 del V1-2	(987)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp 160 del 128-194	(1032)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp140TM	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp140	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp140mut	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp120	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
Consensus	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
		1321	1360
		GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	
gp160	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	
gp160 del V1	(1225)	GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	
gp160 del V2	(1240)	GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	
gp160 del V1-2	(1027)	GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	
gp 160 del 128-194	(1072)	GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	
gp140TM	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	
gp140	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	
gp140mut	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	
gp120	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	
Consensus	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCATCCGGGGCC	

FIG. 66B-7

104 / 131		1400
	1361	
gp160	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCTGTGCTGAC
gp160 del V1	(1265)	AGATCAAGTGCAGCAGCAATATTACCGGCTGTGCTGAC
gp160 del V2	(1280)	AGATCAAGTGCAGCAGCAATATTACCGGCTGTGCTGAC
gp160 del V1-2	(1067)	AGATCAAGTGCAGCAGCAATATTACCGGCTGTGCTGAC
gp 160 del 128-194	(1112)	AGATCAAGTGCAGCAGCAATATTACCGGCTGTGCTGAC
gp140TM	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCTGTGCTGAC
gp140	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCTGTGCTGAC
gp140mut	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCTGTGCTGAC
gp120	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCTGTGCTGAC
Consensus	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCTGTGCTGAC
	1401	
gp160	(1401)	CCGCGACGGCGGCCACCAACAACCGCACCAACGACACC
gp160 del V1	(1305)	CCGCGACGGCGGCCACCAACAACCGCACCAACGACACC
gp160 del V2	(1320)	CCGCGACGGCGGCCACCAACAACCGCACCAACGACACC
gp160 del V1-2	(1107)	CCGCGACGGCGGCCACCAACAACCGCACCAACGACACC
gp 160 del 128-194	(1152)	CCGCGACGGCGGCCACCAACAACCGCACCAACGACACC
gp140TM	(1401)	CCGCGACGGCGGCCACCAACAACCGCACCAACGACACC
gp140	(1401)	CCGCGACGGCGGCCACCAACAACCGCACCAACGACACC
gp140mut	(1401)	CCGCGACGGCGGCCACCAACAACCGCACCAACGACACC
gp120	(1401)	CCGCGACGGCGGCCACCAACAACCGCACCAACGACACC
Consensus	(1401)	CCGCGACGGCGGCCACCAACAACCGCACCAACGACACC
	1441	
gp160	(1441)	GAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAAC
gp160 del V1	(1345)	GAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAAC
gp160 del V2	(1360)	GAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAAC
gp160 del V1-2	(1147)	GAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAAC
gp 160 del 128-194	(1192)	GAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAAC
gp140TM	(1441)	GAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAAC
gp140	(1441)	GAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAAC
gp140mut	(1441)	GAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAAC
gp120	(1441)	GAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAAC
Consensus	(1441)	GAGACCTTCCGCCCGCGCGCGGCAACATGAAGGACAAC
	1481	
gp160	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGGCGCATCGA
gp160 del V1	(1385)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGGCGCATCGA
gp160 del V2	(1400)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGGCGCATCGA
gp160 del V1-2	(1187)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGGCGCATCGA
gp 160 del 128-194	(1232)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGGCGCATCGA
gp140TM	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGGCGCATCGA
gp140	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGGCGCATCGA
gp140mut	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGGCGCATCGA
gp120	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGGCGCATCGA
Consensus	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGGCGCATCGA
	1521	
gp160	(1521)	GCCCTTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG
gp160 del V1	(1425)	GCCCTTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG
gp160 del V2	(1440)	GCCCTTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG
gp160 del V1-2	(1227)	GCCCTTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG
gp 160 del 128-194	(1272)	GCCCTTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG
gp140TM	(1521)	GCCCTTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG
gp140	(1521)	GCCCTTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG
gp140mut	(1521)	GCCCTTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG
gp120	(1521)	GCCCTTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG
Consensus	(1521)	GCCCTTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG

FIG. 66B-8

SUBSTITUTE SHEET (RULE 26)

105 / 131

		1561	1600
gp160	(1561)	GTGCAGCGCGAGAAGCGCGCGCTGGGCCTGGGCGCCCTGT	
gp160 del V1	(1465)	GTGCAGCGCGAGAAGCGCGCGCTGGGCCTGGGCGCCCTGT	
gp160 del V2	(1480)	GTGCAGCGCGAGAAGCGCGCGCTGGGCCTGGGCGCCCTGT	
gp160 del V1-2	(1267)	GTGCAGCGCGAGAAGCGCGCGCTGGGCCTGGGCGCCCTGT	
gp 160 del 128-194	(1312)	GTGCAGCGCGAGAAGCGCGCGCTGGGCCTGGGCGCCCTGT	
gp140TM	(1561)	GTGCAGCGCGAGAAGCGCGCGCTGGGCCTGGGCGCCCTGT	
gp140	(1561)	GTGCAGCGCGAGAAGCGCGCGCTGGGCCTGGGCGCCCTGT	
gp140mut	(1561)	GTGCAGCGCGAGAAGCGCGCTGGGCCTGGGCGCCCTGT	
gp120	(1561)	GTGCAGCGCGAGAAGCGCTAAG-----	
Consensus	(1561)	GTGCAGCGCGAGAAGCGCGCGCTGGGCCTGGGCGCCCTGT	
		1601	1640
gp160	(1601)	TCATCGGCTTC-CTGGGCGCGCGGGAGCACCATTGGGCG	
gp160 del V1	(1505)	TCATCGGCTTC-CTGGGCGCGCGGGAGCACCATTGGGCG	
gp160 del V2	(1520)	TCATCGGCTTC-CTGGGCGCGCGGGAGCACCATTGGGCG	
gp160 del V1-2	(1307)	TCATCGGCTTC-CTGGGCGCGCGGGAGCACCATTGGGCG	
gp 160 del 128-194	(1352)	TCATCGGCTTC-CTGGGCGCGCGGGAGCACCATTGGGCG	
gp140TM	(1601)	TCATCGGCTTC-CTGGGCGCGCGGGAGCACCATTGGGCG	
gp140	(1601)	TCATCGGCTTC-CTGGGCGCGCGGGAGCACCATTGGGCG	
gp140mut	(1601)	TCATCGGCTTC-CTGGGCGCGCGGGAGCACCATTGGGCG	
gp120	(1583)	ATATCGGATCCTCTAGA-----	
Consensus	(1601)	TCATCGGCTTCCTGCGCGCGGGAGCACCATTGGGCG	
		1641	1680
gp160	(1640)	CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp160 del V1	(1544)	CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp160 del V2	(1559)	CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp160 del V1-2	(1346)	CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp 160 del 128-194	(1391)	CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp140TM	(1640)	CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp140	(1640)	CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp140mut	(1640)	CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp120	(1600)	-----	
Consensus	(1641)	CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
		1681	1720
gp160	(1680)	GAGCGGCATCGTGACGAGCAGAAACACCTGCTGCGCGCC	
gp160 del V1	(1584)	GAGCGGCATCGTGACGAGCAGAAACACCTGCTGCGCGCC	
gp160 del V2	(1599)	GAGCGGCATCGTGACGAGCAGAAACACCTGCTGCGCGCC	
gp160 del V1-2	(1386)	GAGCGGCATCGTGACGAGCAGAAACACCTGCTGCGCGCC	
gp 160 del 128-194	(1431)	GAGCGGCATCGTGACGAGCAGAAACACCTGCTGCGCGCC	
gp140TM	(1680)	GAGCGGCATCGTGACGAGCAGAAACACCTGCTGCGCGCC	
gp140	(1680)	GAGCGGCATCGTGACGAGCAGAAACACCTGCTGCGCGCC	
gp140mut	(1680)	GAGCGGCATCGTGACGAGCAGAAACACCTGCTGCGCGCC	
gp120	(1600)	-----	
Consensus	(1681)	GAGCGGCATCGTGACGAGCAGAAACACCTGCTGCGCGCC	
		1721	1760
gp160	(1720)	ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V1	(1624)	ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V2	(1639)	ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V1-2	(1426)	ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp 160 del 128-194	(1471)	ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp140TM	(1720)	ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp140	(1720)	ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp140mut	(1720)	ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp120	(1600)	-----	
Consensus	(1721)	ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	

FIG. 66B-9

106/131

		1761	1800
gp160	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCTGGCCGTGGAGCG	
gp160 del V1	(1664)	GCATCAAGCAGCTGCAGGCCCGCATCTGGCCGTGGAGCG	
gp160 del V2	(1679)	GCATCAAGCAGCTGCAGGCCCGCATCTGGCCGTGGAGCG	
gp160 del V1-2	(1466)	GCATCAAGCAGCTGCAGGCCCGCATCTGGCCGTGGAGCG	
gp 160 del 128-194	(1511)	GCATCAAGCAGCTGCAGGCCCGCATCTGGCCGTGGAGCG	
gp140TM	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCTGGCCGTGGAGCG	
gp140	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCTGGCCGTGGAGCG	
gp140mut	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCTGGCCGTGGAGCG	
gp120	(1600)	-----	
Consensus	(1761)	GCATCAAGCAGCTGCAGGCCCGCATCTGGCCGTGGAGCG	1840
		1801	1840
gp160	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp160 del V1	(1704)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp160 del V2	(1719)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp160 del V1-2	(1506)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp 160 del 128-194	(1551)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp140TM	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp140	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp140mut	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp120	(1600)	-----	
Consensus	(1801)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	1880
		1841	1880
gp160	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp160 del V1	(1744)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp160 del V2	(1759)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp160 del V1-2	(1546)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp 160 del 128-194	(1591)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp140TM	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp140	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp140mut	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp120	(1600)	-----	
Consensus	(1841)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	1920
		1881	1920
gp160	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp160 del V1	(1784)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp160 del V2	(1799)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp160 del V1-2	(1586)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp 160 del 128-194	(1631)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp140TM	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp140	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp140mut	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp120	(1600)	-----	
Consensus	(1881)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	1960
		1921	1960
gp160	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp160 del V1	(1824)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp160 del V2	(1839)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp160 del V1-2	(1626)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp 160 del 128-194	(1671)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp140TM	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp140	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp140mut	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp120	(1600)	-----	
Consensus	(1921)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	

FIG. 66B-10

SUBSTITUTE SHEET (RULE 26)

107 / 131

		1961	2000
gp160	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCGAAGCC	
gp160 del V1	(1864)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCGAAGCC	
gp160 del V2	(1879)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCGAAGCC	
gp160 del V1-2	(1666)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCGAAGCC	
gp 160 del 128-194	(1711)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCGAAGCC	
gp140TM	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCGAAGCC	
gp140	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCGAAGCC	
gp140mut	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCGAAGCC	
gp120	(1600)	-----	
Consensus	(1961)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCGAAGCC	
		2001	2040
gp160	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V1	(1904)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V2	(1919)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V1-2	(1706)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp 160 del 128-194	(1751)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp140TM	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp140	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp140mut	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp120	(1600)	-----	
Consensus	(2001)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
		2041	2080
gp160	(2040)	GTGGGCCAGCCTGTGGAACCTGGTTCGACATCACCAACTGG	
gp160 del V1	(1944)	GTGGGCCAGCCTGTGGAACCTGGTTCGACATCACCAACTGG	
gp160 del V2	(1959)	GTGGGCCAGCCTGTGGAACCTGGTTCGACATCACCAACTGG	
gp160 del V1-2	(1746)	GTGGGCCAGCCTGTGGAACCTGGTTCGACATCACCAACTGG	
gp 160 del 128-194	(1791)	GTGGGCCAGCCTGTGGAACCTGGTTCGACATCACCAACTGG	
gp140TM	(2040)	GTGGGCCAGCCTGTGGAACCTGGTTCGACATCACCAACTGG	
gp140	(2040)	GTGGGCCAGCCTGTGGAACCTGGTTCGACATCACCAACTGG	
gp140mut	(2040)	GTGGGCCAGCCTGTGGAACCTGGTTCGACATCACCAACTGG	
gp120	(1600)	-----	
Consensus	(2041)	GTGGGCCAGCCTGTGGAACCTGGTTCGACATCACCAACTGG	
		2081	2120
gp160	(2080)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp160 del V1	(1984)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp160 del V2	(1999)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp160 del V1-2	(1786)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp 160 del 128-194	(1831)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp140TM	(2080)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp140	(2080)	CTGTGGTACATC-----	
gp140mut	(2080)	CTGTGGTACATC-----	
gp120	(1600)	-----	
Consensus	(2081)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
		2121	2160
gp160	(2120)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCA----	
gp160 del V1	(2024)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCA----	
gp160 del V2	(2039)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCA----	
gp160 del V1-2	(1826)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCA----	
gp 160 del 128-194	(1871)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCA----	
gp140TM	(2120)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCATCGT	
gp140	(2092)	-----	
gp140mut	(2092)	-----	
gp120	(1600)	-----	
Consensus	(2121)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCANNNN	

FIG. 66B-11

108 / 131

		2161	2200
gp160	(2156)	-TCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	
gp160 del V1	(2060)	-TCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	
gp160 del V2	(2075)	-TCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	
gp160 del V1-2	(1862)	-TCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	
gp 160 del 128-194	(1907)	-TCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	
gp140TM	(2160)	GTAAGATATCGGATCCTCTAGA-----	
gp140	(2092)	-TAAGATATCGGATCCTCTAGA-----	
gp140mut	(2092)	-TAAGATATCGGATCCTCTAGA-----	
gp120	(1600)	-----	
Consensus	(2161)	NTCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	2240
		2201	2240
gp160	(2195)	TGCAGACCCGCGCTGCCGCCAGCGCGGCCCGACCGCCC	
gp160 del V1	(2099)	TGCAGACCCGCGCTGCCGCCAGCGCGGCCCGACCGCCC	
gp160 del V2	(2114)	TGCAGACCCGCGCTGCCGCCAGCGCGGCCCGACCGCCC	
gp160 del V1-2	(1901)	TGCAGACCCGCGCTGCCGCCAGCGCGGCCCGACCGCCC	
gp 160 del 128-194	(1946)	TGCAGACCCGCGCTGCCGCCAGCGCGGCCCGACCGCCC	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2201)	TGCAGACCCGCGCTGCCGCCAGCGCGGCCCGACCGCCC	2280
		2241	2280
gp160	(2235)	CGAGGGCATCGAGGAGGAGGGCGCGAGCGCGACCGCGAC	
gp160 del V1	(2139)	CGAGGGCATCGAGGAGGAGGGCGCGAGCGCGACCGCGAC	
gp160 del V2	(2154)	CGAGGGCATCGAGGAGGAGGGCGCGAGCGCGACCGCGAC	
gp160 del V1-2	(1941)	CGAGGGCATCGAGGAGGAGGGCGCGAGCGCGACCGCGAC	
gp 160 del 128-194	(1986)	CGAGGGCATCGAGGAGGAGGGCGCGAGCGCGACCGCGAC	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2241)	CGAGGGCATCGAGGAGGAGGGCGCGAGCGCGACCGCGAC	2320
		2281	2320
gp160	(2275)	CGCAGCAACCGCTGGTGCACGGCTGTCTGGCCCTGATCT	
gp160 del V1	(2179)	CGCAGCAACCGCTGGTGCACGGCTGTCTGGCCCTGATCT	
gp160 del V2	(2194)	CGCAGCAACCGCTGGTGCACGGCTGTCTGGCCCTGATCT	
gp160 del V1-2	(1981)	CGCAGCAACCGCTGGTGCACGGCTGTCTGGCCCTGATCT	
gp 160 del 128-194	(2026)	CGCAGCAACCGCTGGTGCACGGCTGTCTGGCCCTGATCT	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2281)	CGCAGCAACCGCTGGTGCACGGCTGTCTGGCCCTGATCT	2360
		2321	2360
gp160	(2315)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACACCG	
gp160 del V1	(2219)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACACCG	
gp160 del V2	(2234)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACACCG	
gp160 del V1-2	(2021)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACACCG	
gp 160 del 128-194	(2066)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACACCG	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2321)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACACCG	

FIG. 66B-12

109/131		2400
2361		
gp160	(2355)	CCTGCGGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG
gp160 del V1	(2259)	CCTGCGGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG
gp160 del V2	(2274)	CCTGCGGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG
gp160 del V1-2	(2061)	CCTGCGGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG
gp 160 del 128-194	(2106)	CCTGCGGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG
gp140TM	(2182)	-----
gp140	(2113)	-----
gp140mut	(2113)	-----
gp120	(1600)	-----
Consensus	(2361)	CCTGCGGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG
2401		2440
gp160	(2395)	CTGCTGGGCGCGCGCGCTGGGAGGCCCTGAAGTACTGGT
gp160 del V1	(2299)	CTGCTGGGCGCGCGCGCTGGGAGGCCCTGAAGTACTGGT
gp160 del V2	(2314)	CTGCTGGGCGCGCGCGCTGGGAGGCCCTGAAGTACTGGT
gp160 del V1-2	(2101)	CTGCTGGGCGCGCGCGCTGGGAGGCCCTGAAGTACTGGT
gp 160 del 128-194	(2146)	CTGCTGGGCGCGCGCGCTGGGAGGCCCTGAAGTACTGGT
gp140TM	(2182)	-----
gp140	(2113)	-----
gp140mut	(2113)	-----
gp120	(1600)	-----
Consensus	(2401)	CTGCTGGGCGCGCGCGCTGGGAGGCCCTGAAGTACTGGT
2441		2480
gp160	(2435)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
gp160 del V1	(2339)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
gp160 del V2	(2354)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
gp160 del V1-2	(2141)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
gp 160 del 128-194	(2186)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
gp140TM	(2182)	-----
gp140	(2113)	-----
gp140mut	(2113)	-----
gp120	(1600)	-----
Consensus	(2441)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
2481		2520
gp160	(2475)	CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC
gp160 del V1	(2379)	CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC
gp160 del V2	(2394)	CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC
gp160 del V1-2	(2181)	CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC
gp 160 del 128-194	(2226)	CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC
gp140TM	(2182)	-----
gp140	(2113)	-----
gp140mut	(2113)	-----
gp120	(1600)	-----
Consensus	(2481)	CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC
2521		2560
gp160	(2515)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT
gp160 del V1	(2419)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT
gp160 del V2	(2434)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT
gp160 del V1-2	(2221)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT
gp 160 del 128-194	(2266)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT
gp140TM	(2182)	-----
gp140	(2113)	-----
gp140mut	(2113)	-----
gp120	(1600)	-----
Consensus	(2521)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT

FIG. 66B-13

111 / 131

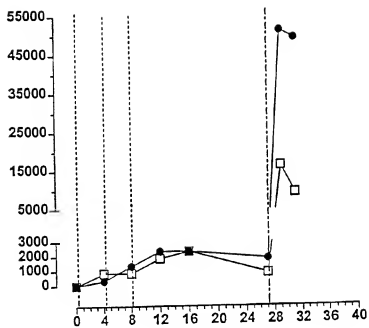


FIG. 67

112 / 131

HIV-1SF2 wt RT (PISPIET-->GIRKVL)
CCCATTAGTCCTATTGAACTGTACCAGTAAATTTAAAGCCAGGAATGGATGGCCCAAAA
GTTAAGCAATGGCCATTGACAGAAGAAAAATTAAGCATTAGTAGAGATATGTACAGAA
ATGGAAAAGGAAGGGAATAATTTCAAAAATTTGGGCTGAAAAATCCATACAATCTCCAGTA
TTTGCTATAAGAAAAAGACAGTACTAAATGGAGAAAACTAGTAGATTTTCAGAGAACTT
AATAAAGAACTCAAGACTTCTGGGAAGTTCAGTTAGGAATACACACCCCGCAGGGTTA
AAAAAGAAAAAATCAGTAACAGTATTGGATGTGGGTGATGCATCTTTTCAGTTCCCTTA
GATAAAGACTTTAGAAAGTATACTGCACTTTACCATACCTAGTATAAACCAATGAGACACCA
GGGATTAGATATCAGTACAATGTGCTGCCACAGGGATGGAAAGGATCACAGCAATATTC
CAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAAACAGAATCCAGACATAGTTATC
TATCAATacatggatgatTTGTATGTAGGATCTGACTTAGAAAAAGGGCAGCATAGAACA
AAAAATAGAGGAACAGAGACAGCATCTGTTGAGGTGGGGATTACCAACCCAGACAAAAAA
CATCAGAAAAGAACCTCCATTCCTTggatgggttatGAACTCCATCTGATAAATGGACA
GTACAGCCTATAATGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAGTTA
GTGGGAAAAATGAAATGGGCAAGTCAGATTTATGCAGGGATTAAAGTAAAGCAGTTATGT
AAACTCCTTAGAGGAACCAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCAGAG
CTAGAACTGGCAGAAAAAGGGAGATTCTAAAGAAGCAAGTACATGAAGTATATTATGAC
CCATCAAAAGACTTAGTAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATATCAA
ATTTATCAAGAGCCATTTAAAAATCTGAAAAACAGGAAAGTATGCAAGGATGAGGGGTGCC
CACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAGTATCCACAGAAAGCATA
GTAATATGGGGAAGATTCTTAAATTTAAACTACCCATACAAAAGGAACATGGGAAGCA
TGGTGGATGGAGTATTGGCAAGCTACCTGGATTCTGAGTGGGAGTTTGTCAATACCCCT
CCCTTAGTGAATTTAGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAACTTTC
TATGTAGATGGGGCAGCTAATAGGGAGACTAAATAGGAAAAGCAGGATATGTTACTGAC
AGAGGAAGACAAAAAGTTGTCTCCATAGCTGACACAACAAATCAGAAGACTGAATTACAA
GCAATTATCTAGCTTTGAGGATTGCGGATTAGAAGTAAACATAGTAACAGACTCACAA
TATGCATTAGGAATCATTTCAAGCACAACAGATAAAGAGTGAATCAGAGTTAGTCAGTCAA
ATAATAGAGCAGTTAATAAAAAAGGAAAGGTCTACCTGGCATGGGTACAGCACACAAA
GGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTCTGGAATCAGGAAAGTACTA

FIG. 68

(SEQ ID NO:77)

113 / 131

GagProtMod.SF2 (GP1)

GTCTGACGCCACCATGGGCGCCCGGCCAGCGTGCTGAGCGGCGGCGAGCTGGACAAGTGG
GAGAAGATCCGCCTGCGCCCCGGCGGGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGG
GCCAGCCGCGAGCTGGAGCGCTTCGCGGTGAACCCCGGCTGCTGGAGACCAGCGAGGGC
TGCCGCCAGATCCTGGGCCAGCTGCAGCCAGCGCTGCAGACCCGCGAGCGAGGAGCTGCGC
AGCCTGTACAACAACCGTGGCCACCCCTGTACTGCGTGCACCCAGCGCATCGAGCTCAAGGAC
ACCAAGGAGGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAAGTCCAAGAAGAAGGGCCAG
CAGGCGCCGCCGCCGCCGCCGCCACCAGGCAACAGCAGCCAGGTGAGCCAGAAGTACCCCATC
GTGCAGAACCTGCAGGGCCAGATGGTGCACCCAGGCCATCAGCCCCCGCACCCCTGAACGCC
TGGGTGAAGGTGGTGGAGGAGAAGGCCCTTCAGCCCCGAGGTGATCCCATGTTTCAGCGCC
CTGAGCGAGGGCGCCACCCCCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCCAC
CAGGCGCCCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCGCGCCGAGTGGGACCGC
GTGACCCCGTGCACCGCGGCCCATCGCCCCCGGCCAGATGCGCGAGCCCCGCGGCAGC
GACATCGCCGCGCACCCAGCACCCCTGCAGGAGCAGATCGGCTGGATGACCAACAACCC
CCCATCCCCGTGGCGAGATCTACAAGCGTGGATCATCCTGGGCTGAACAAGATCGT
CGGATGTACAGCCCCACAGCATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCGCG
GACTACGTGGACCGCTTCTACAAGACCTGCGCGCTGAGCAGGCCAGCCAGGACGTGAAG
AACTGGATGACCGAGACCCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAGACCATCCTG
AAGGCTCTCGGCCCCCGCGGCCACCCCTGGAGGAGATGATGACCGCTGCCAGGCGCTGGG
GGCCCCGGCCACAAGGCCCCGCGTGTGGCCGAGGCGATGAGCCAGGTGACGAACCCGGG
ACCATCATGATGCAGCGCGGCAACTTCCGCAACGAGCGGAAGACCGTCAAGTGTCTCAAC
TGCGGCAAGGAGGSCCACACCGCCAGGAAGTGCAGCGCCCCCGCAAGAAGGGCTGCTGG
CGCTGCGGCGCGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTA
GGGAAGATCTGGCCTTCTACAAGGGAAGGCCAGGGAATTTCTTCAGAGCAGACCAGAG
CCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAG
AAGCAGGAGCCGATAGACAAGGAAGTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGC
AACGACCCCTCGTACAGTAAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCG
GCGCCGACGACACCGTGTGGAGGAGATGAACCTGCCCGCAAGTGAAGCCCCAAGATGA
TCGCGGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCT
GCGGCCACAAGGCCATCGGCACCGTGTGGTGGGCCCCACCCCGTGAACATCATCGGCC
GCAACCTGCTGACCCAGATCGGCTGCACCCCTGAAGTTCCCCATCAGCCCCATCGAGACGG
TGCCCGTGAAGCTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCCTGTAAG
AATTC

FIG. 69
(SEQ ID NO:78)

114 / 131

GagProtMod.SF2 (GP2)

GTCGACGCCACCATGGGCGCCCGCCAGCGTGCTGAGCGGCGCGAGCTGGACAAGTGG
GAGAAGATCCGCTTGGCCCCGGCGGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGG
GCCAGCCGCGAGCTGGAGCGCTTTCGCCGTGAACCCCGCCTGCTGGAGACCAGCGAGGGC
TGCCGCCAGATCCTGGGCCAGCTGCAGCCAGCCTTCAGACCGGACGCGAGGAGCTGCCG
AGCCTGTACAACACCGTGGCCACCCTGTACTGCGTGACACGCGCATCGACGTCAAGGAC
ACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAAGTCCAAGAAGAAGGCCACG
CAGGCCGCGCGCGCCGCGGCACCGGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATC
GTGCAGAACCTGCAGGGCCAGATGGTGCACCAGGCCATCAGCCCCCGCACCTGAAACGCC
TGGGTGAAGGTGGTGGAGGAGAAGGCCCTTCAGCCCCGAGGTGATCCCCATGTTTCAGCGCC
CTGAGCGAGGGCGCCACCCCCCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCCAC
CAGSGCCGCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCCGCGAGTGGGACCGC
GTGCACCCCGTGCACGCCGCGCCCATCGCCCCCGGCCAGATGCGCGAGCCCCGCGGCAGC
GACATCGCCGGCACCCAGCACCCCTGCAGGAGCAGATCGGCTGGATGACCAACAACCCC
CCCCATCCCGTGGGCGAGATCTACAAGCGGTGGATCATCCTGGGCTGAACAAGATCGTG
CGGATGTACAGCCCCCAGCATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCGCG
GACTACGTGGACCGCTTCTACAAGACCCCTGCGCGCTGAGCAGGCCAGCCAGSACGTGAAG
AACTGGATGACCGAGACCCCTGCTGGTGCAGAACGCCAACCCGACTGCAAGACCATCCTG
AAGGCTCTCGGCCCCGCGGCCACCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGC
GGCCCCGGCCACAAGGCCCGCGTGTGGCCGAGGCGATGAGCCAGGTGACGAACCCGGCG
ACCATCATGATGCAGCGCGGCAACTCCGCAACCAGCGGAAGACCGTCAAGTGCTTCAAC
TGCGGCAAGGAGGGCCACACCGCCAGGAATGCCGCGCCCCCGCAAGAAGGGGCTGCTGG
CGCTGCGGCGCGCAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTTTTTA
GGGAAGATCTGGCTTCTACAAGGGAAGGCCAGGGAATTTCTTCAGAGCAGACCAGAG
CCAACAGCCCCCACCAGAAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACCTCCCTCTCAG
AAGCAGGAGCCGATAGACAAGGAAGTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGC
AACGACCCCTCGTCACAGTAAGGATCGGGGGCAACTCAAGGAAGCGCTGCTCGATACAG
GAGCAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAATGGAACCAAAAAATGA
TAGGGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCT
GTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAA
GAAATCTGTGACCCAGATCGGCTGCACCTTGAACCTTCCCATCAGCCCTATTGAGACGG
TGCCCGTGAAGTTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGTAAG
AATTC

FIG. 70

(SEQ ID NO:79)

115 / 131

FS(+)_ProtInact_RTpt_YM

GCGGCCGCGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGGA
AGATCTTGGCCTTCTACAAGGGAAGGCCAGGGAATTTCTTCAGAGCAGACCAGAGCCAA
CAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGAGGAGAAAACAACTCCCTCTCAGAAGC
AGGAGCCGATAGACAAGGAACCTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACG
ACCCCTCGTCAACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGC
AGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAACCAAAAATGATAGG
GGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTGG
ACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAA
TCTGTTGACCCAGATCGGCTGCACCTTGAACCTCCCCATCAGCCCTATTGAGACGGTGCC
CGTGAAGTTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGGA
GAAGATCAAGGCCCTGTTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAA
GATCGGCCCCGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAGAAGGACAGCAC
CAAGTGGCGCAAGCTGTTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGA
GGTGCAGCTGGGCATCCCCACCCCGCGGCTGAAGAAGAAGAAGAGCGTGACCGTGCT
GGACGTGGGCGACGCCCTACTTCAGCGTGCCCTGGACAAGGACTTCCGCAAGTACACCGC
CTTCACCATCCCCAGCATCAACAACGAGACCCCGGCATCCGCTACCAGTACAACGTGCT
GCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCAGAGCAGCATGACCAAGATCCTGGA
GCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCTGTACGTGGGCAG
CGACCTGGAGATCGGCCAGCACCCGACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCG
CTGGGGCTTCACCAACCCCGACAAGAAGCACCAAGGAGCCCCCTTCTGTGGATGGG
CTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCAGAGAAGGACAG
CTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTA
CGCCGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGCACCAAGGCCCTGACCGA
GGTGATCCCCCTGACCGAGGAGGCGGAGCTGGAGCTGGCCGAGAACC GCGAGATCCTGAA
GGAGCCCGTGACAGAGGTGTACTACGACCCAGCAAGGACCTGTTGGCCGAGATCCAGAA
GCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGAC
CGGCAAGTACGCCCGCATGCGCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGC
CGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCT

FIG. 71A

(SEQ ID NO:80)

116 / 131

GCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGAT
CCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAA
GGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAA
GCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGA
CACCAACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCT
GGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCCGA
CAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGT
GTACCTGGCCTGGGTGCCCCGCCACAAGGGCATCGGCGGCCAACGAGCAGGTGGACAAGCT
GGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCCTGAACGGCATCGATGGCGGCATCGTGAT
CTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAGAAGCT
TCCCGGGGCTAGCACCGGTGAATTC

FIG. 71B

(SEQ ID NO:80)

117 / 131

FS(+)_ProtInact_RTOpt_YMWM

GC GGCCCGGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTTATAGGGA
AGATCTGGCCTTCTCAACAAGGGAAGGCCAGGGAATTTCTTCAGAGCAGACCAGAGCCAA
CAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACCTCCCTCTCAGAAGC
AGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCTCAGATCACTCTTTGGCAACG
ACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGC
AGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAATGATAGG
GGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTGG
ACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAA
TCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGCC
CGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGGA
GAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAA
GATCGGCCCCGAGAACCCTACAACACCCCCGTGTTCCGCATCAAGAAGAAGGACAGCAC
CAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGA
GGTGCAGCTGGGCATCCCCACCCCGCGCTGAAGAAGAAGAAGAGCGTGACCGTGCT
GGACGTGGGCGACGCTACTTCAGCGTGCCCCGTGGACAAGGACTTCCGCAAGTACACCGC
CTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACAGTACAACGTGCT
GCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCAGAGCAGCATGACCAAGATCCTGGA
GCCCTTCCGCAAGCAAAACCCGACATCGTGATCTACAGGCCCCCTGTAGCTGGGCAG
CGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCG
CTGGGGCTTCAACACCCCCGACAAGAAGCACCAAGAGAGCCCCCTTCTGCCCATCGA
GCTGCACCCGACAAGTGGACCGTGCGAGCCCATCATGTGCTGCCGAGAAGGACAGCTGGAC
CGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGG
CATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCCGAGGTGAT
CCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCC
CGTGACCGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGG
CCAGGGCCAGTGGACCTACAGATCTACCAAGAGCCCTTCAAGAACCTGAAGACCGGCAA
GTACGCCCCGATCGCGCGGCCCAACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCA
GAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAGCTGCCCAT

FIG. 72A

(SEQ ID NO:81)

118 / 131

CCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCGA
GTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCC
CATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGG
CAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCAC
CAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGT
GAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCCGACAAGAG
CGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCT
GGCCTGGGTGCCCCGCCACAAGGGCATCGGCCGCCAACGAGCAGGTGGACAAGCTGGTGAG
CGCCGGCATCCGCAAGGTGCTGTTCTCTGAACGGCATCGATGGCGGCATCGTGATCTACCA
GTACATGGACGACCTGTACGTGGGCAGCGCGGCCCTAGGATCGATTAAAAGCTTCCCCG
GGCTAGCACCGGTGAATTC

FIG. 72B

(SEQ ID NO:81)

119 / 131

FS(-)_ProtMod_RTpt_YM

GCGGCCGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG
AGGACCTGGCCTTCTGTCAGGGCAAGGCCCGCGAGTTTCAGCAGCGAGCAGACCCGCGCCA
ACAGCCCCACCCGCGCGAGCTGCAGGTGTGGGCGGCGAGAACACAGCCTGAGCGAGG
CCGGCGCGGACCGCCAGGGCACCCTGAGCTTCAACTTCCCCCAGATCACCTGTGGCAGC
GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG
CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAAGCCCAAGATGATCG
GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG
GCCACAAGGCCATCGGCACCCTGCTGGTGGGCCCCACCCCGTGAACATCATCGGCCGCA
ACCTGTGACCCAGATCGGCTGCACCTGAACTTCCCCATCAGCCCCCATCGAGACGGTGC
CCGTGAAGCTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCCTGACCAGG
AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA
AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTGCCATCAAGAAGAAGGACAGCA
CCAAGTGGCGCAAGCTGGTGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG
AGGTGCAGCTGGGCATCCCCACCCCGCGGCGCTGAAGAAGAAGAGAGCGTGACCGTGC
TGGACGTGGGCGACGCTACTTCAGCGTGCCCCCTGGACAAGGACTTCCGCAAGTACACCG
CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGCGCATCCGCTACCAGTACAACGTGC
TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCAGAGCAGCATGACCAAGATCCTGG
AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACAGGCCCCCTGTACGTGGGCA
GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTGC
GCTGGGCTTCACCAACCCCGACAGAAGCACCAGAAGGAGCCCCCTTCTGTGGATGG
GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA
GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAGTGGGCCAGCCAGATCT
ACGCCGCGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCAACCAAGGCCCTGACCG
AGGTGATCCCCCTGACCGAGGAGGCGAGCTGGAGCTGGCCGAGAACCAGGAGATCCTGA
AGGAGCCCGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATCCAGA
AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA
CCGGCAAGTACGCCCGCATGCGCGGCGCCACCAACGACGTGAAGCAGCTGACCGGAGG
CCGTGCAGAAGGTGAGCACCAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC

FIG. 73A

(SEQ ID NO:82)

120 / 131

TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA
TCCCCGAGTGGGAGTTCGTGAACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA
AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA
AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG
ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC
TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCCCG
ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG
TGTACCTGGCCTGGGTGCCCCGCCACAAGGGCATCGGCGGCCAACGAGCAGGTGGACAAGC
TGGTGAGCGCCGCATCCGCAAGGTGCTGTTCTGAAACGGCATCGATGGCGGCATCGTGA
TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGCGGCCCTAGGATCGATTAAAAGC
TTCCCGGGGCTAGCACCGGTGAATTC

FIG. 73B

(SEQ ID NO:82)

121 / 131

FS(-)_ProtMod_RTpt_YMWM

GCGGCCGCGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGGG
AGGACCTGGCCTTCTTGCAGGGCAAGGCCCGCGAGTTTCAGCAGCGAGCAGACCCGCGCCA
ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG
CCGGCGCCGACCCGAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCTGTGGCAGC
GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGTGTGTCGACACCGGCG
CCGACGACACCGTGTGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG
GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG
GCCACAAGGCCATCGGCACCGTGTGGTGGGCCCCACCCCGTGAAATCATCGGCCGCA
ACCTGTGACCCAGATCGGCTGCACCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC
CCGTGAAGCTGAAGCCGGGGATGGACGCGCCCCAAGGTCAAGCAGTGGCCCCGTGACCGAGG
AGAAGATCAAGGCCCTGTGGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA
AGATCGGCCCGGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAGAAGGACAGCA
CCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG
AGGTGACAGCTGGGCATCCCCACCCCGCGGCGCTGAAGAAGAAGAAGAGCGTGACCGTGC
TGGACGTGGGCGACGCTACTTTCAGCGTGCCTTGGACAAGGACTTCCGCAAGTACACCG
CCTTCACCATCCCCAGCATCAACAACGAGACCCCGGCATCCGCTACCAAGTACAACGTGC
TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCTGTG
AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAAGGCCCCCTGTACGTGGGCA
GCGACCTGGAGATCGGCCAGCACCGCACCAGATCGAGGAGCTGCGCCAGCACCTGCTGC
GCTGGGGCTTCAACACCCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCTGCCCATCG
AGCTGCACCCCGACAAGTGGACCGTGCGAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA
CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG
GCATCAAGGTGAAGCAGCTGTGAAGCTGTGCGCGGCACCAAGGCCCTGACCGAGGTGA
TCCCCCTGACCGAGGAGGCCGAGCTGAGAGCTGGCCGAGAACCAGGAGATCTGAAGGAGC
CCGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG
GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA
AGTACGCCCGCATGCGCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC
AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA

FIG. 74A

(SEQ ID NO:83)

122 / 131

TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGCGAGGCCACCTGGATCCCCG
AGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGTACCAGCTGGAGAAGGAGC
CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG
GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA
CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG
TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCCCGACAAGA
GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC
TGGCCTGGGTGCCCGCCACAAGGGCATCGGCGGCCAACGAGCAGGTGGACAAGCTGGTGA
GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC
AGTACATGACGACCTGTACGTGGGCAGCGCGGCCCTAGGATCGATTAAAAGCTTCCCG
GGGCTAGCACCGGTGAATTC

FIG. 74B

(SEQ ID NO:83)

123 / 131

FS(-)_ProtMod_RTOpt(+)

GCGGCCGCGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG
AGGACCTGGCCTTCTGCAAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA
ACAGCCCCACCCGCCGAGCTGCAAGGTGTGGGCGCGGAGAACACAGCCTGAGCGAGG
CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCTGTGGCAGC
GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGCG
CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGCAAGTGAAGCCCAAGATGATCG
GCGGGATCGGGGGCTTCAACAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG
GCCACAAGGCCATCGGCACCCGTGCTGGTGGGCCCCACCCCGTGAAACATCATCGGCGCGCA
ACCTGTGTGACCCAGATCGGCTGCACCCCTGAACCTTCCCCCATCAGCCCCATCGAGACGCTGC
CCGTGAAGCTGAAGCCGCGGATGGACGGCCCCAAGSTCAAGCAGTGGCCCCCTGACCGAGG
AGAAGATCAAGGCCCTTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAGATCAGCA
AGATCGGCCCGGAGAACCCTTACAACACCCCGTGTTCGCCATCAAGAAGAAGGACAGCA
CCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG
AGGTGCAGCTGGGCATCCCCACCCCGCGGCGCTGAAGAAGAAGAAGAGCGTGACCGTGC
TGGACGTGGGCGACGCCCTACTTCAGCGTGCCCTGGACAAGGACTTCCGCAAGTACACCG
CCTTACCATCCCCAGCATCAACAACGAGACCCCGGCATCCGCTACAGTACAACGTGC
TGCCCCAGGGCTGGAAGGGCGAGCCCCGCATCTTCCAGAGCAGCATGACCAAGATCCTGG
AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG
TGGGCGAGCGACCTGGAGATCGGCGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC
TGCTGCGCTGGGGCTTCAACACCCCGACAAGAAGCACAGAAGGAGCCCCCTTCTCTGT
GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGACGCCCATCATGTGCCCCGAGA
AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGTTGGGCAAGCTGAACTGGGCCAGCC
AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCC
TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA
TTCTGAAGAGCCCTGTCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGA
TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACAGATCTACCAGSAGCCCTTCAAGAACC
TGAAGACCGGCAAGTACGCCCGCATGCGCGGCCGCCACCAACGACGTGAAGCAGCTGA
CCGAGGCCCTGTCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT
TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA
CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACACAG
TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGACGCGCCCGCCCAACCGCG

FIG. 75A

(SEQ ID NO:84)

124 / 131

AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA
TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA
GCGGCCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC
AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG
AGAAGGTGTACCTGGCCTGGGTGCCCCGCCCAAGGGCATCGGCGGCAACGAGCAGGTGG
ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCTGAACGGCATCGATGGCGGCA
TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT
AAAAGCTTCCCCGGGGCTAGCACCGGTGAATTC

FIG. 75B
(SEQ ID NO:84)

125 / 131

Tat_wt_SF162 (wildtype)

ATGGAGCCAGTAGATCCTAGATTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAGA
 CTGCTTGTAACAAATTGCTATTGTAAGGAGTTGCTTTCATTGCCAAGTTTGTTCATAAC
 AAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGAGCTCCT
 CCAGACAGTGAGGTTTCATCAAGTTTCTCTACCAAGCAACCCGCTTCCAGCCCCAAGG
 GGACCCGACAGGCCCCGAAGGAATCGAAGAAGAAGGTGGAGAGAGACAGAGACAGA
 TCCAGTCCATTAG

FIG. 76

(SEQ ID NO:85)

Tat_SF162

MEPVDPRLEPWKHPGSQPKTACTNICYCKKCCFHCQVCFITKGLISYGRKKRRQRRRAPDSE
 VHQVSLPKQPASQPGDPTGPKESKKKVERETETDPVH

FIG. 77

(SEQ ID NO:86)

Tat_SF162_opt

ATGGAGCCCGTGGACCCCCGCCTGGAGCCCTGGAAGCACCCCGGCAGCCAGCCCCAAGAC
 CGCCTGCACCAACTGCTACTGCAAGAAGTGCTGCTTCCACTGCCAGGTGTGCTTCATCACC
 AAGGGCCTGGGCATCAGCTACGGCCGCAAGAAGCGCCGCCAGCGCCGCCGCCCCCCC
 CGACAGCGAGGTGCACCAAGGTGAGCCTGCCCAAGCAGCCCGCCAGCCAGCCCCAGGGCG
 ACCCCACCGGCCCCAAGGAGAGCAAGAAGAAGGTGGAGCGCGAGACCAGACCCGACCCC
 TGCACTAG

FIG. 78

(SEQ ID NO:87)

Tat_Cys22_SF162_opt

ATGGAGCCCGTGGACCCCCGCCTGGAGCCCTGGAAGCACCCCGGCAGCCAGCCCCAAGAC
 CGCCTGCACCAACTGCTACTGCAAGAAGTGCTGCTTCCACTGCCAGGTGTGCTTCATCACC
 AAGGGCCTGGGCATCAGCTACGGCCGCAAGAAGCGCCGCCAGCGCCGCCGCCCCCCC
 GACAGCGAGGTGCACCAAGGTGAGCCTGCCCAAGCAGCCCGCCAGCCAGCCCCAGGGCG
 CCCCACCGGCCCCAAGGAGAGCAAGAAGAAGGTGGAGCGCGAGACCAGACCCGACCCC
 TGCACTAG

FIG. 79

(SEQ ID NO:88)

126 / 131

Alignment GagMod vs GP1_GP2

	1	10	20	30	40	50	60	76	Section 1
GagMod_SF2	(1)	1							
GagProtMod_SF2 (GP1)	(1)	ATGGGGCCCCGGCCAGCGCTGCTGAGCGCGCGGAGCTGGCAAGTGGGAGAGATCGCCTGGCCCGCCGGCGCA							
GagProtMod_SF2 (GP2)	(1)	ATGGGGCCCCGGCCAGCGCTGCTGAGCGCGCGGAGCTGGCAAGTGGGAGAGATCGCCTGGCCCGCCGGCGCA							
Consensus	(1)	ATGGGGCCCCGGCCAGCGCTGCTGAGCGCGCGGAGCTGGCAAGTGGGAGAGATCGCCTGGCCCGCCGGCGCA							
									Section 2
GagMod_SF2	(77)	77	90	100	110	120	130	140	152
GagProtMod_SF2 (GP1)	(77)	AGAGAGAGTACAAAGCTGAAGCACATCTGTGGGCGAGCGGAGCTGGAGGCGCTTGGCGTGAACCCCGGCGCTGCT							
GagProtMod_SF2 (GP2)	(77)	AGAGAGAGTACAAAGCTGAAGCACATCTGTGGGCGAGCGGAGCTGGAGGCGCTTGGCGTGAACCCCGGCGCTGCT							
Consensus	(77)	AGAGAGAGTACAAAGCTGAAGCACATCTGTGGGCGAGCGGAGCTGGAGGCGCTTGGCGTGAACCCCGGCGCTGCT							
									Section 3
GagMod_SF2	(153)	153	160	170	180	190	200	210	228
GagProtMod_SF2 (GP1)	(153)	GGAGACGACGAGGGGCTGCGGCCAGATCTCTGGGCGAGCTGCAAGCCAGCTGCAGCCGGGAGCGAGGAGCTGGCG							
GagProtMod_SF2 (GP2)	(153)	GGAGACGACGAGGGGCTGCGGCCAGATCTCTGGGCGAGCTGCAAGCCAGCTGCAGCCGGGAGCGAGGAGCTGGCG							
Consensus	(153)	GGAGACGACGAGGGGCTGCGGCCAGATCTCTGGGCGAGCTGCAAGCCAGCTGCAGCCGGGAGCGAGGAGCTGGCG							
									Section 4
GagMod_SF2	(229)	229	240	250	260	270	280	290	304
GagProtMod_SF2 (GP1)	(229)	AGCCTGTACACACACCGTGGCCACCTGTACTGCTGCACCGCATCGAGCTCAAGGACACCAAGAGAGGCCCTGG							
GagProtMod_SF2 (GP2)	(229)	AGCCTGTACACACACCGTGGCCACCTGTACTGCTGCACCGCATCGAGCTCAAGGACACCAAGAGAGGCCCTGG							
Consensus	(229)	AGCCTGTACACACACCGTGGCCACCTGTACTGCTGCACCGCATCGAGCTCAAGGACACCAAGAGAGGCCCTGG							
									Section 5
GagMod_SF2	(305)	305	310	320	330	340	350	360	370
GagProtMod_SF2 (GP1)	(305)	AGAAGATCGAGGAGGACGAGACAACTCCAGAAAGAGGCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCAACAG							
GagProtMod_SF2 (GP2)	(305)	AGAAGATCGAGGAGGACGAGACAACTCCAGAAAGAGGCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCAACAG							
Consensus	(305)	AGAAGATCGAGGAGGACGAGACAACTCCAGAAAGAGGCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCAACAG							

FIG. 80A

128 / 131

Alignment GagMod vs GP1_GP2

	761	770	780	790	800	810	820	836	Section 11
GagMod.SF2	(761)								
GagProtMod.SF2 (GP1)	(761)	ACACCC	CCCATC	CCCGTGGG	GAGATCTAC	AGCGGCTG	ATCTCCTGG	CTGAACAGATCTG	CGGATGTA
GagProtMod.SF2 (GP2)	(761)	ACACCC	CCCATC	CCCGTGGG	GAGATCTAC	AGCGGCTG	ATCTCCTGG	CTGAACAGATCTG	CGGATGTA
Consensus	(761)	ACACCC	CCCATC	CCCGTGGG	GAGATCTAC	AGCGGCTG	ATCTCCTGG	CTGAACAGATCTG	CGGATGTA
	837	850	860	870	880	890	900	912	Section 12
GagMod.SF2	(837)								
GagProtMod.SF2 (GP1)	(837)	CAGCCCC	ACACGATCT	GGACATCCG	CAGGCGCCCA	AGAGCCCTTCC	CGCACTAC	TGTGGACCGCTT	TACAAG
GagProtMod.SF2 (GP2)	(837)	CAGCCCC	ACACGATCT	GGACATCCG	CAGGCGCCCA	AGAGCCCTTCC	CGCACTAC	TGTGGACCGCTT	TACAAG
Consensus	(837)	CAGCCCC	ACACGATCT	GGACATCCG	CAGGCGCCCA	AGAGCCCTTCC	CGCACTAC	TGTGGACCGCTT	TACAAG
	913	920	930	940	950	960	970	988	Section 13
GagMod.SF2	(913)								
GagProtMod.SF2 (GP1)	(913)	ACCC	TGGCGCTG	AGCAGCGCC	AGGACGTGA	AGACTGGATG	ACCGAGACCTT	GTGTGCGAAGCCCAACC	
GagProtMod.SF2 (GP2)	(913)	ACCC	TGGCGCTG	AGCAGCGCC	AGGACGTGA	AGACTGGATG	ACCGAGACCTT	GTGTGCGAAGCCCAACC	
Consensus	(913)	ACCC	TGGCGCTG	AGCAGCGCC	AGGACGTGA	AGACTGGATG	ACCGAGACCTT	GTGTGCGAAGCCCAACC	
	989	1000	1010	1020	1030	1040	1050	1064	Section 14
GagMod.SF2	(989)								
GagProtMod.SF2 (GP1)	(989)	CCGACT	CGAAGACAT	CTTGAGG	CTCTCG	CCCGCGCCACCTT	GGAGAGATGAT	CACCGCTGCC	AGGCGGT
GagProtMod.SF2 (GP2)	(989)	CCGACT	CGAAGACAT	CTTGAGG	CTCTCG	CCCGCGCGCCACCTT	GGAGAGATGAT	CACCGCTGCC	AGGCGGT
Consensus	(989)	CCGACT	CGAAGACAT	CTTGAGG	CTCTCG	CCCGCGCGCCACCTT	GGAGAGATGAT	CACCGCTGCC	AGGCGGT
	1065	1070	1080	1090	1100	1110	1120	1140	Section 15
GagMod.SF2	(1065)								
GagProtMod.SF2 (GP1)	(1065)	GGGCGG	CCCCGCGC	ACAGGCCCCG	CGCTGTG	CGCCAGCGCAT	GAGCCAGGTGAC	GAACCCCGGACCATCATGATG	
GagProtMod.SF2 (GP2)	(1065)	GGGCGG	CCCCGCGC	ACAGGCCCCG	CGCTGTG	CGCCAGCGCAT	GAGCCAGGTGAC	GAACCCCGGACCATCATGATG	
Consensus	(1065)	GGGCGG	CCCCGCGC	ACAGGCCCCG	CGCTGTG	CGCCAGCGCAT	GAGCCAGGTGAC	GAACCCCGGACCATCATGATG	

FIG. 80C

131/ 131

TataminoSF162.opt

ATGGAGCCCGTGGACCCCGCCTGGAGCCCTGGAAGCACC'CCGGCAGCCAGCCCAA
GACCGCCCTGGACCAACTGCTACTGCAAGAAGTGTGCTTCCACTGCGCAGGTGTGCTT
CATCACCAAGGCGCTGGGCATCAGCTACGGCGCGAAGAAGCGCGCCGCCAGCGCGCG

FIG. 81
(SEQ ID NO:89)

TaL_Cys22_SF162

MEPVDPRLEPWKHPGSPKTAGTNCYKKKCFHCQVCFITKGLGISYGRKKRRRRAPPDSE
VHQVSLPKQPASQPQGDPTGPKESKKVERETETDPVHZ

FIG. 82
(SEQ ID NO:90)

SEQUENCE LISTING

<110> Chiron Corporation

<120> IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
OF VIRUS-LIKE PARTICLES

<130> 1621.100

<140>

<141>

<160> 90

<170> Patentin Ver. 2.0

<210> 1

<211> 1509

<212> DNA

<213> Human immunodeficiency virus

<400> 1

```

atgggtcgca gagcgctcggg attaagcggg ggagaattag ataaatggga aaaaattcgg 60
ttaaggccag ggggaaagaa aaaatataag ttaaaacata tagtatgggc aagcaggggag 120
ctagaacgat tcgcagtcgaa tctctggcctg ttagaacatc cagaaggctg cagacaaata 180
ttgggacagc tacagccatc ccttcagaca ggatcagaag aacttagatc attatataat 240
acagtacgaa cctcttatgt tgtacatcaa aggtatagatg taaaagacac caaggaaagt 300
ttagagaaga tagagaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360
gcagctggca caggaacacag cagccaggtc agccaaaatt accttatagt gcagaacctt 420
caggggcaaa tggtagatca ggccatatca cctagaactt taatgcatg ggtaaagtaa 480
gtagaagaaa aggcctttcag cccagaagta ataccatgt ttccagcatt atcagaagga 540
gccacccccc aagattttaa cccatgctta aacacagtgg ggggacatca agcagccatg 600
caaatgttaa aagagactat caatgaggaa gctgcagaat gggatagagt gcatccagtg 660
catgcagggc ctattgcacc aggcctaatg agagaaccaa ggggaagtga catagcagga 720
actactagta ccttcaggga acaaatagga tggatgacaa ataaccacc tatccagta 780
ggagaaatct ataaaagatg gataatcctg ggattaaata aaatagtaag aatgtatagc 840
cctaccagca ttctggacat aagacaagga ccaagggaac ccttagaga ttagttagac 900
cggttctata aaactctaaag agccgaacaa gcttcacagg atgtaaaaaa ttgtagtaca 960
gaaacctgtg tggcccaaaa tgcacaacca gattgtaaga ctattttaaa agcatgtgga 1020
ccagcagcta cactagaaga aatgatgaca gcatgtcagg gagtgggggg acccgcccat 1080
aaagcaagag ttttggctga agccatgagc caagtaacaa atccagctaa cataatgatg 1140
cagagaggca attttaggaa ccaaaagaaa actgttaagt gtttcaattg tggcaaaaga 1200
gggacacacc ccaaaaattg cagggccctc agggaaaaag gctgttgga atgtggaagg 1260
gaaggacacc aaatgaanga ttgcactgag agacaggcta attttttagg gaagatctgg 1320
ccttcttaca aggggaagcc aggggaattt ctccagagca gaccagagcc aacagcccca 1380
ccagagaaga gcttcaggtt tggggaggag aaaaacactc cctctcagaa gcaggagccg 1440
atagacaagg aactgtatcc tttaacttcc ctccagatcac tctttggcaa cgacccctcg 1500
tcacaataa

```

1509

<210> 2

<211> 1845

<212> DNA

<213> Human immunodeficiency virus

<400> 2

```

atgggtcgca gagcgctcggg attaagcggg ggagaattag ataaatggga aaaaattcgg 60
ttaaggccag ggggaaagaa aaaatataag ttaaaacata tagtatgggc aagcaggggag 120
ctagaacgat tcgcagtcgaa tctctggcctg ttagaacatc cagaaggctg cagacaaata 180

```

ttgggacagc	tacagccatc	ccttcagaca	ggatcagaag	aacttagatc	attatataat	240
acagttagca	cctctattg	tgtacatcaa	aggatagatg	taaaagacac	caaggaaagt	300
ttagagaaga	tagaggaaga	gcaaaacaaa	agtaagaaaa	aggcacagca	agcagcagct	360
gcagctcgca	caggaaacag	cagccagagtc	agccaaaactt	accctatagt	gcagaaacct	420
caggggcaaa	tggtacatca	ggccatcatca	cctagaactt	taaatgcatg	ggtaaaagta	480
gtagaagaat	aggcttttcag	cccagaagta	ataccatagt	tttcagcatt	atcagaagga	540
gccacccac	agatntttaa	caccatgcta	aacacagtgg	ggggacatca	agcgcagctg	600
caaatgttaa	agagagctat	caatgaggaa	gctgcagaat	gggatagagt	gcatacagtg	660
catgcagggc	ctattgcacc	aggccaaatg	agagaaacca	ggggaaagtga	agcagcagga	720
actactagta	cccttcagga	acaaatagga	tggtatgaca	ataatccacc	tatcccaagta	780
ggagaaatct	ataaaagatg	gataatcctg	ggattaaata	aaatagtaag	aattgtatgc	840
ctaccacgca	ttctggacat	aagacaagga	ccaaaggaaac	cctttagaga	ttatgtagac	900
cggttctata	aaactctaa	agccgaacaa	gcttcacagg	atgtaaaaaa	ttggatgaca	960
gaaaccttgt	tgttccaaaa	tgcaaaccca	gattgttaaga	ctatttttaa	agcattggga	1020
ccagcagcta	cactagaaga	aatgatgaca	gcattgcagg	gagtgggggg	accggccat	1080
aaagcaagag	ttttggctga	agccatgagc	caagttaacaa	atccagctaa	cataatgatg	1140
caagagagca	atttttaggaa	ccaaagaaga	actgttaagt	gtttcaattg	tggcaaaaga	1200
gggcacatag	ccaaaaattg	cagggccctt	aggaataaag	gctgttggag	atgtgggaag	1260
gaaggacac	aaatgaaaga	ttgcactgag	agacaggcta	atttttttagg	gaagatctgg	1320
cctctctaca	aggggaagcc	agggaaattt	cttcagagca	gaccagagcc	aacagcccca	1380
ccagaagaga	gcttcagggt	tggggaggag	aaaacaactc	cctctcagaa	gcaggagccg	1440
atagacaagg	aactgtatcc	tttaacttcc	ctcagatcac	ttttggcaa	gcacctctgc	1500
tcacaataag	ataggggggg	caactaaagg	aagctctatt	agatacagga	gcagatgata	1560
cagttataga	agaaatgaat	ttgccaggaa	aatggaaacc	aaaaatgata	gggggaattg	1620
gaggttttat	caaatgaaga	cagtacgata	agatacctgt	agaaatctgt	ggacataaag	1680
ctatagttac	agttattagta	ggacctacac	ctgtcaacat	aattggaaaga	aactcgttga	1740
ctcagatgg	ttgtacttta	aatttcccca	ttagtctcat	tgaaactcta	ccagtaaaat	1800
ttaagccagg	aatggatggc	ccaaaagtta	agcaatggcc	attga		1845

<210> 3

<211> 4313

<212> DNA

<213> Human immunodeficiency virus

<400> 3

atgggtgcga	gagcgtcgg	attaagcggg	ggagaattag	ataaatggga	aaaaattcgg	60
ttaaagccag	ggggaaagaa	aaaaataag	ttaaaacata	tagtatgggc	aagcaggagg	120
ctagaacgat	tcgcagtcac	tctgcgctgt	ttagaacaat	cagaaggctg	cagacaaata	180
ttgggacagc	tacagccatc	ccttcagaca	ggatcagaag	aacttagatc	attatataat	240
acagtgcgaa	ccctctattg	tgtacatcaa	aggatagatg	taaaagacac	caaggaaagt	300
ttagagaaga	tagaggaaga	gcaaaacaaa	agtaagaaaa	aggcacagca	cagcagagct	360
gcagctcgca	caggaaacag	cagccagagtc	agccaaaact	accctatagt	gcagaaacct	420
caggggcaaa	tggtacatca	ggccatcatca	cctagaactt	taaatgcatg	ggtaaaagta	480
gtagaagaat	aggcttttcag	cccagaagta	ataccatagt	tttcagcatt	atcagaagga	540
gccacccac	agattttaa	caccatgcta	aacacagtgg	ggggacatca	agcgcagctg	600
caaatgttaa	agagagctat	caatgaggaa	gctgcagaat	gggatagagt	gcatacagtg	660
catgcagggc	ctattgcacc	aggccaaatg	agagaaacca	ggggaaagtga	catagcagga	720
actactagta	cccttcagga	acaaatagga	tggtatgaca	ataatccacc	tatcccaagta	780
ggagaaatct	ataaaagatg	gataatcctg	ggattaaata	aaatagtaag	aattgtatgc	840
cctaccagca	ttctggacat	aagacaagga	ccaaaggaaac	cctttagaga	ttatgtagac	900
cggttctata	aaactctaa	agccgaacaa	gcttcacagg	atgtaaaaaa	ttggatgaca	960
gaaaccttgt	tgttccaaaa	tgcaaaccca	gattgttaaga	ctatttttaa	agcattggga	1020
ccagcagcta	cactagaaga	aatgatgaca	gcattgcagg	gagtgggggg	accggccat	1080
aaagcaagag	ttttggctga	agccatgagc	caagttaacaa	atccagctaa	cataatgatg	1140
cagagaagca	atttttaggaa	ccaaagaaag	actgttaagt	gtttcaattg	tggcaaaaga	1200
gggcacatag	ccaaaaattg	cagggccctt	aggaataaag	gctgttggag	atgtgggaag	1260
gaaggacac	aaatgaagaa	ttgcactgag	agacaggcta	atttttttagg	gaagatctgg	1320
cctctctaca	aggggaagcc	agggaaattt	cttcagagca	gaccagagcc	aacagcccca	1380
ccagaagaga	gcttcagggt	tggggaggag	aaaacaactc	cctctcagaa	gcaggagccg	1440

```

atagacacagg aactgtatcc ttttaacttcc ctcagatcac tctttggcga cgacccctcg 1500
tcacataaag gatagggggg caactaaagg aagctctatt agatcacaga gcgagatgata 1560
cagtagtaga agaaatgaat ttgccaggaa aatggaaacc aaaaatgata ggggggaattg 1620
caggttttat caaagtaaga cagtagcatc agatacctgt agaaatctgt ggcataaaag 1680
ctataggtag agtattagta ggacctacac ctgtcaacat aattggaaaga aatctgttga 1740
ctcagattgg ttgtacttta aatttcccca ttagtctcat tgaactgta ccagtaaaat 1800
taaaagccagg aatggtggc ccaaaagtta agcaatggcc attgacagaa gaagaaaaataa 1860
aagcattagt agagatatgt acagaaatgt aaaaaggagg gaaaatttca aaaaattgggc 1920
ctgaataatcc atacaatact ccagttattg ctataaagaa aaaagacagt actaatggga 1980
gaaaactagt agattttcaga gaacttaata aaagaactca agactctgtg gaagtccaagt 2040
taggaataacc acaccccgca gggttaaaaa agaaaaaac agtaacagta ttgggtgtgg 2100
gtgatgcata cttttcagtt cccttagata aagactttag aaagtatact gcatttacca 2160
tacctagtat aaacaatgag acaccaggga ttagatatca gtacaatgtg ctgccacagg 2220
gtggaaaagg atccaccaga atattccaaa gtacatgac aaaaatttta gagcctttta 2280
gaaaacagaa tccagacata gttatctatc aatacatgga tgaattgtat gtaggatctg 2340
acttgaanaa agggcgagcat agaacaataa tagaggaaact gagacagcat ctgttgaggt 2400
ggggatttac cacaccagac aaaaacatc agaaagacc tccattctct tggatgggtt 2460
atgaactcca tctctgataa tggacagtag agcctataat gctgccagaa aaagacagct 2520
ggactgtcaa tgacatacag aagtttagtg gaaaaattgaa ttgggcaagt cagatttatgt 2580
ggggatttaa agtaaacgag ttatgtaaac tcttagagg aaccaaagca ctaacagaaag 2640
taataccact aacagaaaga gcagagctag aactggcaga aaacaggagg attctaaaag 2700
aaccagtaca tgaagtatat tatgacctat caaaagactt agtagcagaa atacagaagc 2760
agggggcaagg ccaatggaca tatcaaatat tcaagagacc atttaaaaaa ctgaaaacag 2820
gaaagtatgc aaggatgagg ggtgccaca ctaatgatgt aaaaacagta acagaggcag 2880
tgcaaaaagt atccacagaa agcatagtaa tatggggaaa gattcctaaa tttaaacata 2940
ccatacaaaa ggaacacatg gaagcatggt ggatggagta ttggcgaagt acctggatg 3000
ctgagtggga ttgtgtcaat accctccct tagtgaattt atggtaccag ttgagaaaag 3060
aacccatagt aggagcagaa actttctatg tagatggggc agetaatagg gagactaaat 3120
taggaaaagg aggatattgt actgacagag gaagacaaa agtgtctctc atagctgaca 3180
caacaaatca gaagactgaa ttacaagcaa ttcattagc ttgacaggat tcgggattag 3240
aagttaaatc agtaacagac tcacaaatag cattaggaat cattcaagca caaccagata 3300
agagtgaatc agagttagtc agtcaataaa tagagcagtt aataaaaaag gaaaaggctc 3360
acctggcatg ggtaccagca cacaaaggaa ttggaggaaa tgaaccaagta gataaattag 3420
tcagtctctg aatcaggaaa gtactatttt tgaatggaa agataaggcc caagaagaac 3480
atgaaaaata tcacagtaat tggagagcaa tggctagata aatgtcagct aaaaaggaga gccatgcatg 3540
tagcaaaaaga aatagtagcc agctgtgata aatgtcagct aaaaaggaga gccatgcatg 3600
gacacataga ctgtagtcca ggaatatggc aactagattg tacacatcta gaagaaaaaa 3660
ttatctctgt agcagttcat gttagccagt gatataaga agcagaagtt attccagcag 3720
agacagggca ggaacacaga tattttctct taaaattgag aggaagatgg ccagtaaaaa 3780
catacatcac agacaatggc agcaatttca ccagtactac gggttaaggcc gcctgtgtgt 3840
gggcagggat agcaggagaa ttgggcattc cctacaatcc ccaaagtcaa ggagagttag 3900
aatctatgaa taatgaatta aagaaaatta taggacaggt aagagatcag gctgaacacc 3960
ttgaacagc agtaaaaatg gcagttattc tccacaatt tttaaaagaaa ggggggattg 4020
ggggatcacag tgcaggggaa agaattagat acataatagc aacagacata caaactaaag 4080
aactcaaaaa gccaatttca aaaaattcaa attttcgggt ttattacagg gacaacaaag 4140
atcccccttt gaaaggacca gcaaaagctc tctggaaaag tgaaggggca gtagttaact 4200
aagataatag tgacataaaa gtatggccaa gaagaaaagc aaaaatcatt agggattatg 4260
gaaaacagat ggcaggtgat gattgtgtgg caagtacaga ggatgaggt tag 4313

```

<210> 4

<211> 1515

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
HIV-Gag

<400> 4

```

gccaccatgg gcgcccgccg cagcgtgctg agcggcgccg agctggacaa gtgggagaag 60
atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatcgt gtgggccagg 120
gcgcagctgg agcgccttcgc cgtgaacccc ggccctgctgg agaccagcga gggctgcgcg 180
cagatcctgg gccagctgca gccacgcctg cagaccggca gcgaggagct gcgcagctg 240
tacaacaccc tggccaccct gtactgcgtg caccagcgca tcgacgtcaa ggaacacca 300
gaggccctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
gcgcgcgcgc ccggcaccgg caacagcagc caggtgagcc atcagccccc gaccctgaa cgccctgggt 420
aacctgcagg gccagatggt gcaccaggcc atcagccccc gaccctgaa cgccctgggt 480
aaggtgggtg aggagaaggc ctcagcccc gaggtgatcc ccatgttcag ccgctgagc 540
gaggggccca cccccaggga cctgaacacg atgttgaaca ccgtggggcg ccaccaggcc 600
gccatgcaga tgctgaagga gaccatcaac gaggaggcgc ccgagtggga ccgctgac 660
cccgctgcag ccggccccat cgcccccgcc cagatgcgcg agcccccgcg cagcgacatc 720
gcgcgcacca ccagcacctc gcaggagcag atcgctgga tgaccaacaa ccccccatc 780
cccgctggcg agatctacaa ccggtggatc atcctggggc tgaaacaagt cgtgcggatg 840
tacagcccca ccagcatcct ggacatccgc caggggccca agggagccct ccgcgactac 900
gtggagcgcg tctacaagac cctgcgcgct gaggcagcca gccaggacgt gaagaactcg 960
atgaccgaga cctgtggtg gcagaacgac aaccccgact gcaagacat cctgaaggct 1020
ctcgcccccg cggccacccct ggaggagatg atgaccgct gccaggcggt gggcgggccc 1080
ggccacaagg ccgcgctgct ggccgaggcg atgaccaggg tgacgaaccc gcgcacatc 1140
atgatgcagc ggcccaactt ccgcaacagg cggaaagacc tcaagtgtct caactgcggc 1200
aaggaggggc acaccgcgag gaactgcgcg gcccccgcga agaaggctg ctggcgctgc 1260
ggcgccgagg gccaccagat gaaggactgc accgagcgcc agggccaact cctgggcaag 1320
atctggccca gctacaaggc ccgccccggc aacttctcgc agagcccgcc cgagccccc 1380
gccccccccc agggagaggt cgtcttcggc gaggagaaga cccccccag ccagaagcag 1440
ggcaccatcg caaggagct gtaccacctg accagctcgc gcagcctgtt cggcaacgac 1500
cccagagcgc agtaa

```

<210> 5

<211> 1853

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

HIV-Gag-protease

<400> 5

```

gccaccatgg gcgcccgccg cagcgtgctg agcggcgccg agctggacaa gtgggagaag 60
atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatcgt gtgggccagg 120
gcgcagctgg agcgccttcgc cgtgaacccc ggccctgctgg agaccagcga gggctgcgcg 180
cagatcctgg gccagctgca gccacgcctg cagaccggca gcgaggagct gcgcagctg 240
tacaacaccc tggccaccct gtactgcgtg caccagcgca tcgacgtcaa ggaacacca 300
gaggccctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
gcgcgcgcgc ccggcaccgg caacagcagc caggtgagcc agaactaccc catcgtgca 420
aacctgcagg gccagatggt gcaccaggcc atcagccccc gaccctgaa cgccctgggt 480
aaggtgggtg aggagaaggc ctcagcccc gaggtgatcc ccatgttcag ccgctgagc 540
gaggggccca cccccaggga cctgaacacg atgttgaaca ccggtggggc ccaccaggcc 600
gccatgcaga tgctgaagga gaccatcaac gaggaggcgc ccgagtggga ccgctgac 660
cccgtycagc ccggccccat cgcccccgcc cagatgcgcg agcccccgcg cagcgacatc 720
gcgcgcacca ccagcacctc gcaggagcag atcgctgga tgaccaacaa ccccccatc 780
cccgtycggc agatctacaa ccggtggatc atcctgggccc tgaaacaagt cgtcgagatg 840
tacagcccca ccagcatcct ggacatccgc caggggccca agggagccct ccgcgactac 900
gtggagcgcg tctacaagac cctgcgcgct gaggcagcca gccaggagct gaagaactcg 960
atgaccgaga cctgtggtg gcagaacgac aaccccgact gcaagacat cctgaaggct 1020
ctcgcccccg cggccacccct ggaggagatg atgaccgct gccaggcggt gggcgggccc 1080
ggccacaagg ccgcgctgct ggccgaggcg atgaccagg tgacgaaccc ggcgaccatc 1140
atgatgcagc gcggcaactt ccgcaacagg cggaaagacc tcaagtgtct caactgcggc 1200
aaggaggggc acaccgcgag gaactgcgcg gcccccgcga agaagggtg cttggcgctg 1260
ggccgcaag gacacccaat gaaagattgc actgagagac aggcataatt ttaggggaag 1320

```

atctgcgctt	cttacaagg	aaggccagg	aattttcttc	agagcagacc	agagccaaca	1380
gcccccacag	aagagagctt	caggtttggg	gaggagaaaa	caactccctc	tcagaagcag	1440
gagccgatag	acaaggaaat	gtatccttta	acttccctca	gatcactctt	tgccaacgac	1500
ccctcgtcac	agtaaggatc	ggcgccagcc	tcaaggagge	gctgctcgac	accggcgccg	1560
acgacacgtg	gctggaggag	atgaacctgc	ccggcaagtg	gaagcccaag	atgatccggc	1620
ggatcggggg	cttcatcaag	gtgcggcagt	acgaccagat	ccccgtggag	atctcgccgc	1680
acaaggccat	cggcacccgt	ctggtggggc	ccacccccgt	gaacatcatc	ggccgcaacc	1740
tgtgacacca	gatcggtctg	accctgaact	tccccatcag	ccccatcgag	acgggtgccg	1800
tgaaagtga	gccgggggat	gacggcccca	aggtcaagca	gtggcccccgt	taa	1853

<210> 6

<211> 4319

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
HIV-Gag-polymerase

<400> 6

gccaccatgg	ggcggcgccg	cagcgtgctg	agcggcgggc	agctgggaca	gtggggagaag	60
atccgctctg	gccccggcgg	caagaagaag	tacaagctga	agcacatcgt	gtggggcagc	120
cgcgagctgg	agcgccttgc	cgtgaacccc	ggcctgctgg	agaccagcga	gggctgcccg	180
cagatctcgg	gccagctgca	gcccagccgt	cagacccggca	gcgaggagct	gcgcagcctg	240
tacacacacc	tgcccaccct	gtactgcgtg	caccagcgca	tcgacgtcaa	ggacaccaag	300
gaggccctgg	agaagatcga	ggaggagcag	aacaagtcca	agaagaaggc	ccagcaggcg	360
gcggccgcgc	ccggcacccg	caacagcagc	caggtgagcc	agaactacc	catcgtgcag	420
aacctgcagg	cccgatgggt	gcaccaggcc	atcagccccc	gcacctgaa	cgctgggtg	480
aaggtgggtg	aggagaaggc	cttcagcccc	gaggtgatcc	ccatgttacc	cgccctgagc	540
gaggggcgca	ccccccagga	cctgaacacg	atgttgaaac	ccgtggcgcg	ccaccaggcc	600
cccatcgaga	tctgaagga	gacctaacac	gaggaggccg	ccgagtggga	ccgcgtgcac	660
cccgctgcag	ccggccccc	cgccccccgc	cagatgcgcg	agcccccgcg	cagcggaatc	720
gcggcgacca	ccagcaccct	gcaggagcag	atcggttgga	tgaccaacaa	cccccccatc	780
cccgctggcg	agatctacaa	gcggtggatc	atcctgggcc	tgaaacaagat	cgtgcgggat	840
tacagcccca	ccagcatctc	ggacatccgc	cagggcccca	aggagccctt	cccggaactac	900
gtggagccct	tctacaagac	cctgcgcgct	gagcaggcca	gccaggacgt	gaagaacttg	960
atgaccgaga	ccttgcgtgt	gcagaacgcc	aaccocgaat	gcaagaccat	ctcgaaggct	1020
ctcggccccc	cgccaccctt	ggaggagatg	atgaccgctt	gccaggggct	ggggcgcccc	1080
ggccacaagg	cccgcgtgct	ggccgaggcg	atgagccagg	tgacgaaccc	ggcgaccatc	1140
atgatcgacg	ggggcaactt	ccgcaaccag	cgggaagacc	tcaagtgtct	caactcggcg	1200
aaggagggcc	acggcccgag	gaactgcgcg	gcccccgca	agaagggtcg	ctggcgctgc	1260
ggcgccgaag	gacacaaaat	gaagatttgc	actgagagac	aggctaat	tttaggggaag	1320
atctggcctt	cctacaaggg	aaggccaggg	aattttcttc	agagcagacc	agagccaaca	1380
gcgccaccag	aagagagctt	caggtttggg	gaggagaaaa	caactccctc	tcagaagcag	1440
gagccgatag	caagggaact	gtatccttta	acttccctca	gatcaactct	tgccaacgac	1500
ccctcgtcac	agtaaggatc	ggcgcccgag	tcaaggaggc	gctgctcgac	accggcgccg	1560
acgacacgtg	ctggaggagg	atgaacctgc	ccggcaagtg	gaagcccaag	atgatcggcg	1620
ggatcggggg	cttcatcaag	gtgcggcagt	acgaccagat	ccccgtggag	atctcgccgc	1680
acaagggtcg	cggcacccgt	ctggtggggc	ccacccccgt	gaacatcatc	ggccgcaaac	1740
tgtgacacca	gatcggtctg	accctgaact	tccccatcag	ccccatcgag	acgggtgccg	1800
tgaaagtga	gccgggggat	gacggcccca	aggtcaagca	gtggcccccgt	accgaggaga	1860
agatcaaggc	cgtggtggag	atctgcaacc	agatggagaa	ggaggggcag	atcagcaaga	1920
tcggcccgca	gaacctctac	aacaccccgc	tggttcgcat	caagaagaag	gacagcacca	1980
agtggcgcaa	gctggtggac	ttccgcgagc	tgaacaagcg	caccacggag	ttctggggag	2040
tgcagctggg	catccccacc	cccgcgcggc	tgaagaagaa	gaagagcgtg	accgtgctgg	2100
actggggcga	cgctacttgc	agcgtgcccc	tggacaagga	cttcgcgaag	tacacccgct	2160
tcaccatccc	cagcatcaac	aacgagaccc	ccggctcccg	ctaccagtat	aaogtctctg	2220
cccagggtcg	gaagggcagc	cccgcacatc	tcagagcagc	catgaccaag	atcctggagc	2280
ccttcgcgaa	gcagacaccc	gacatcgtga	tctaccagta	catggacgac	ctgtacgtgg	2340


```

gcagcgacct ggagatcggc cagcaccgca ccaagatcga ggagctgcgc cagcacctgc 2400
tgcgctgggg cttcaccacc cccgacaaga agcaccacga ggagcccccc ttccctgtga 2460
tgggctacga gctgcacccc gacaagtgga ccgtgcagcc catcatgctg cccgagagaag 2520
acagctggac cgtgaacgac atccagaagc tggtagggcaa gctgaactgg gccagccaga 2580
tctacgccgg catcaaggtg aagcagctgt gcaagctgct gcgcggcacc aaggccctga 2640
ccgagtgtag cccctcgacc gaggaggccg agctggagct ggccgagaac gcgcgagatcc 2700
tgaaggagcc cgtgcacgag gtgtactacg accccagcaa ggacctgggt gccgagatcc 2760
agaagcaggg ccaggggccag tggacctacc agatctacca ggagcccttc aagaacctga 2820
agaccgtgca gaaggtgagc accgagagca tcgtgatctg gggcaagatc cccaagtcca 2880
agctgcccat ccagaaggag acctgggagg cctgggtgga ggagtaactg caggccacct 2940
ggatccccga gtgggagttc gtgaacaccc ccccccctgt gaagctgtgg taccaactgg 3000
aagaaggagcc catcgtgggc gccgagacct tctacgtgga cggcgccgcc aaccggcgag 3060
ccaagctggg caaggccggc taccgtgacg accggcgccg ccagaagggt gtgagcatcg 3120
ccgacaccac caaccagaag accgagctgc agggcatcca cctggccctg caggacagcg 3180
gccttgaggt gaacatcgtg accgacagcc agtcacgccct gggaatcatc caggcccagc 3240
ccgacaagag cgaagcgag ctggtgagcc agatcatcga cagctgatc aagaaggaga 3300
aggtgtacct ggccctgggtg cccgcccaaa agggcatcgg cggcaacgag caggtggaca 3360
agctggtgag cgcgggcata cgcaaggtgc tgttccctga cggcatcgac aagcccgagg 3420
aggagacaga gaagtaccac agcaactggc ggcgcatggc caggcacttc aagctgcccc 3480
ccgtggtggc caaggagatc gtggccagct gcgacaagtg ccagctgaag ggcgaggcca 3540
tgacgggcca ggtggactgc agcccggca tctggcagct ggactgacc cactctggag 3600
cgaagatcat cctggtgggc gtgcacgtgg ccagcggcta catcgaggcc gagggtgactc 3660
ccgcccagag accgctact accgctact tctgtcgtaa cgtggccggc gcctggccgg 3720
tgaagacct caacaccgac aacggcagca acttcaccag caccacccgtg aaggccgact 3780
gctgtggggc cggatcaag caggagtctg gcatccctca caaccccag agccaggcg 3840
tggtggagag catgaacaac gaggctgaaga agatcatcgg ccaggtgcgc gaccaggccg 3900
agcacctgaa gcagaccgtg cagatggccc tgttcatcca caacttcaa gcacaaggcg 4020
gcacgtgggg ctacagcgcc ggcgagcgca tcgtggacat catcgccacc gacatccaga 4080
ccaaggagct cgagaagcag atcaccaaga tccagaactc ccgctgtgac taccgagaca 4140
acaaggaccc cctgtggaag ggcgccgcca agctctgtg gaaggggcag ggcgcgctgg 4200
tgatccagga caacagcgac atcaaggtgg tgcccgcgc caaggccaag atcatccgcg 4260
actacggcaa ccagatggcc ggcgacgact cgtgggccag ccgcagagac gaggactag 4319

```

<210> 7

<211> 2031

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
HIV-Gag/HCV-core fusion polypeptide

<400> 7

```

gccacatcag gcgccccgcg cagcgtgctg agcggcgggc agctggacaa gtgggagaag 60
atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatcgt gtggggcagc 120
cgcgagctgg agcgtctcgc cgtgaacccc ggctgctgag agaccagcga gggctgcgcg 180
cagatcctgg gccagctgca gccacgctg cagaccggca gcgagagagt gcgcagcctg 240
tacaacacgc tggccacctc gtactcgctg caccagcgca tcgacgtcaa gcgaaccaag 300
gaggccctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcg 360
gcgcgcggcg ccggcaccgg caacagcagc caggtgagcc agaactcccc catctgtcag 420
aaactcgagg gccagatggt gaccacggcg atcagccccc gacacctgaa gcgctgggtg 480
aaggtgtggg agggagaagc cttcagcccc gaggtgatcc ccatgttcag cgcctcgagt 540
gagggtgcga ccccccagga cctgaacacg atgttgaaac ccgtggggcg ccaccaggcg 600
ggcgcagaga tgcgtgaagg gacctcaac gaggaggcgg ccgagtgga cccggtgcac 660
cccgctgacg ccggccccc cgcctccggc cagatgcgcy agcccccgcg ccgagacatc 720
gccggcacca cagcacacct gcaggagcag atcggctgga tgaccaacaa ccccccatc 780
cctggtggcg agatctacaa gcggtggatc atcctggggc tgaacaagat cgtgcgagct 840
tacagcccca ccagcatctc ggacatccgc caggggccca aggagccctt ccgcgactac 900

```

```

gtggaccgct tctacaagac cctgcgcgct gagcaggcca gccaggacgt gaagaactgg 960
atccagcgaga cctcgtggtt gcagacgcgc aaccgccact cctgaaggct 1020
ctcgcccccgc cgccgaccct ggaggagatg atgaccgcct gccaggcgct gggcgccccc 1080
ggccacaagg ccgcgctgct ggccgaggcg atgagccacc tgacgaaccc ggcacccatc 1140
atgatgcagc gcggcaactt ccgcaaccag cgggaagacg tcaagtgcct caactcgccg 1200
aaggaggccc acaccgcgag gaactgcgcg gcccccgcga agaaggggctg ctggcgctgc 1260
ggccgcgcgag gccaccagat gaaggactgc accgagcgcc aggcccaact cctggggcaag 1320
atctggccca gctacaaggg ccgcgcccgcc aacttctcct agagccgccc cgagcccaacc 1380
gcccccccgc aggagagctt ccgcttcggc gagggagaaga ccaccccagc ccagaagcag 1440
gagcccatcg actaaggagct gtaccacctg accagcctgc gcagcctgtt cggcaacgac 1500
ccagcagccc agtcgacgaa tctctaaact caaagaaaaa acaaacgtta caccacacct 1560
cgcccaacag acgtcaagtt cccgggtgct ggtcagatcg ttggtggagt ttaacttgtt 1620
ccgcgcaggg gccctagatt ggggtgcgc gcgacgagaa agactccga gcggtcgcaa 1680
cctcgaggta gacgtcagcc tatcccaag cgtcgtcgcc ccgagggcag gacctgggct 1740
cagccggggt acccttgccc cctctatggc aatgagggct gcgggtgggc gggatggctc 1800
ctgtctcccc gtggtctcgc gctcagctgc ggccccacg acccccggcg taggtcgccg 1860
aatttgggta aggtctcga tacctctacg tgcggtctcg ccgacctcat ggggtacata 1920
ccgctcgtcg gcgcccctct tggaggcgct gccaggcccc tggcgcatcg cgtccggggt 1980
ctggaagacg gcgtgaacta tgcaacaggg aaccttctg gtgtctetta g 2031

```

<210> 8

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

HIV-Gag/HCV-Core fusion polypeptide

<400> 8

```

atgggtgcga gagcgtcggt attaagcggg ggagaattag ataaattggga aaaaattcgg 60
ttaaggccag ggggaaagaa aaaatataag ttaaaacata tagtatggcg aagcaggggct 120
ctagaacgat tcgcagtcac tctctggcctg ttagaacatc cagaaggctg cgacacaaat 180
ttgggacagc tacagccatc ccttcagaca ggatcagaag aacttagatc attatataat 240
acagtacgaa cctctattg tgtacatcaa aggatagatg taaaagacac caaggaaagt 300
ttagagaaga tagaggaaga gcaaaaacaa agtaagaaaa aggcacagca agcagcagct 360
gcagcttgca caggaaacag cagccaggtc agccaaaatt accctatagt gcagaacct 420
caggggcaca ttgtacatca ggccatatca cctagaactt taaatgcatt ggttaaaagt 480
gtagaagaaa aggtttttag ccagaaagta ataccatgt tttcagcatc atcagaagga 540
gcaacccac acgatttaaa caccatgcta aacacagtg ggggacatca agcagccatg 600
caaatgttaa aagagactat caatgaggaa gctgcagaat gggatagagt gcatccagtg 660
catcgagggc ctactgcacc agggccaaatg agagaaccaa ggggaagtga catagcagga 720
actactagta ccttcaggga acaaatagga tggatgacaa ataatccacc tatcccgata 780
ccagaaactc ataaaaagatg gataatcctg ggattaaata aaatagtaag aatgtatagc 840
ggtaccagca ttctggacat aagacaagga ccaaaaggaa cctttagaga ttatgtagac 900
cggtctctata aaactctaag agccgaacaa gcttcacagg atgtaaaaaa ttggatgaca 960
gaacacctgt ttgtccaaaa tgcaaaccca gattgtaaga ctattttaaa agcatctggga 1020
ccagcagcta cactagaaga aatgatgaca gcattgcagg gagtgggggg acccgcccat 1080
aaagcaagag ttttggctga agccatgagc caagtacaaa atccagctaa cataatgatg 1140
cagagaagga attttaggaa ccaaaagaa actgttaagt gtttcaattg tggcaagaaa 1200
gggcacatag ccaaaaattg cagggccctc aggaaaaagg gctgttggag atgtggaagg 1260
gaagcacacc aaatgaagaa tgcacttag agacaggcta attttttagg gaagatctgg 1320
cttctctaca agggaaaggcc agggaatttt cttcagagca gaccagacc caacgcccc 1380
ccagaagaga gcttcaggtt tggggaggag aaacaacct cctctcagaa gcaggagccg 1440
atagacaagg aactgtatcc tttaacttcc ctcagatcac tctttggcaa gcaccctcg 1500
tcacagtcca cgaatcctaa acctcaaaaa aaaaacaaac gtaaacccaa cgtgcctcca 1560
caggacgtca agttcccggg tggcggtcag atcgttggtg gagtttactt gtgctcgccg 1620
agggggcccta gatgggggt gcgcgcgagc agaaagactt ccgagcggtc gcaacctcga 1680
ggttagacgtc agctatccc caaggctcgt cggcccgagg gcaggacctg ggctcagccc 1740

```

```

gggtaccctt ggccccctta tggcaatgag ggctgcgggt gggcgggatg gctcctctet 1800
ccccgtggct ctccggctag ctggggcccc acagaccccc ggcgtaggtc gcgcaatttg 1860
ggtaaaggtca tcgataccct tacgtgcggc ttccgcgacc tcatggggta cataccgctc 1920
gtcggcgccc ctcttggagg cgctgcgaag gccctggcgc atggcgctcg ggttctggaa 1980
gacggcgtag actatgcaac agggaaacct cctggttgct cttag 2025

```

<210> 9
 <211> 1268
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: synthetic Gag
 common region

```

<400> 9
gccaccatgg gcgccccgc cagcgtgctg agcggcggcg agctggacaa gtgggagaag 60
atccgctctg gccccggcgg caagaagaag tacaagctga agcacatcgt gtggggcagc 120
cgcgagctgg atcggtctgc cgtgaacccc ggctgctgag agaccagcga gggctgcgcg 180
cagatcctgg gccagctgca gccacgctg cagacgggca gcgaggagct gcgcagcctg 240
tacaaacaccg tggccacctt gtactgcgtg caccagcgca tcgacgtcaa ggacaccaag 300
gaggcctctg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
gcgcgcgcgc cgggcacggg caacagcagc caggtgagcc agaactaccc catcgtgcag 420
aacctgcagg gcagatggtt gcaccaggcc atcagccccc gcacctgaa gcgctgggtg 480
aagtggttgg aggagaaggc cttcagcccc gagggtgatcc ccatgttcag gcgctgagc 540
gaggggcgca cccccaggga cctgaacacg atgttgaaca cgtggggcgg ccaccaggcc 600
gccatgcaga tgctgaagga gaccatcaac gaggaggccg ccgagtggga ccgctgcac 660
ccgctgcacg ccggccccat cgcccccggc cagatgcgcg agcccccggg cagcgacatc 720
gcgcgcacaa ccagccacct gcaggagcag atcggtgga tgaccaacaa cccccccatc 780
ccgctggggc agatctacaa gcggtggatc atctggggc tgacaagat cgtgcggatg 840
tacagcccca ccagctcctt ggacatccgc cagggcccca aggagccctt ccgcgactac 900
ctggaccgct tctacaagac cctgcgcgct gagcaggcca gccaggacgt gaagaactgg 960
atgaccgaga cctgtctggt gcagaaagcc aacccgact gcagaccat cctgaaggct 1020
ctcggccccc cgcccaacct ggaggagatg atgaccgct gccaggcgct ggcgcggccc 1080
ggccacaagg ccgcgctggt gcccgaggcg atgaccgagg tgacgaaccc ggccagccatc 1140
atgatgcagc gcggcaactt ccgcaaccag cggaagaccg tcaagtgcct caactgcggc 1200
aaggaggggc acacggccag gaactgcgcg gccccccgca agaagggtg ctggcgctgc 1260
ggcgcgca 1268

```

<210> 10
 <211> 20
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: HIV-Gag
 peptide p7G

```

<400> 10
Gly Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile Asn Glu
1          5          10          15
Glu Ala Ala Glu
20

```

<210> 11
 <211> 30
 <212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer GAG5

<400> 11
aagaattcca tgggtgcgag agcgtcggtta 30

<210> 12
<211> 30
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer
p55-SAL3

<400> 12
attcgtcgac tgtgacgagg ggtcgttgcc 30

<210> 13
<211> 34
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer
CORESALS

<400> 13
atttgcgac gaatccctaaa cctcaaagaa aaac 34

<210> 14
<211> 30
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer 173CORE

<400> 14
tattggatcc taagagcaac caggaagggtt c 31

<210> 15
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer MS65

<400> 15
cgaccatcat ggatgcagcg c 21

<210> 16
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: primer MS66

 <400> 16
 aggattcgtc gagtcgtgc tggggtcgtt 30

 <210> 17
 <211> 26
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: primer XPANXNF

 <400> 17
 gcacgtgggc ccggcgccctc tagagc 26

 <210> 18
 <211> 26
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: primer XPANXNR

 <400> 18
 gctctagagg cgccgggccc acgtgc 26

 <210> 19
 <211> 20
 <212> PRT
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: HIV p55 Gag
 Major Homology Region

 <400> 19
 Asp Ile Arg Gln Gly Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg
 1 5 10 15

 Phe Tyr Lys Thr
 20

 <210> 20
 <211> 60
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: synthetic p55
 Gag Major Homology Region

 <400> 20
 gacatccgcc agggccccaaggagcccttc cgcgactacg tggaccgctt ctacaagacc 60

 <210> 21
 <211> 15

<212> PRT
<213> Human immunodeficiency virus

<400> 21
Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg
1 5 10 15

<210> 22
<211> 5
<212> PRT
<213> Human immunodeficiency virus

<400> 22
Lys Ala Lys Arg Arg
1 5

<210> 23
<211> 4
<212> PRT
<213> Human immunodeficiency virus

<400> 23
Arg Glu Lys Arg
1

<210> 24
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: aa of
mut7.SF162 cleavage site

<400> 24
Ala Pro Thr Lys Ala Ile Ser Ser Val Val Gln Ser Glu Lys Ser
1 5 10 15

<210> 25
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: aa of
mut8.SF162 cleavage site

<400> 25
Ala Pro Thr Ile Ala Ile Ser Ser Val Val Gln Ser Glu Lys Ser
1 5 10 15

<210> 26
<211> 15
<212> PRT

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: aa of
mut.SF162 cleavage site

<400> 26
Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Ser
1 5 10 15

<210> 27
<211> 15
<212> PRT
<213> Human immunodeficiency virus

<220>
<223> Description of Artificial Sequence: aa of native
cleavage site in US4

<400> 27
Ala Pro Thr Gln Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg
1 5 10 15

<210> 28
<211> 5
<212> PRT
<213> Human immunodeficiency virus

<220>
<223> Description of Artificial Sequence: aa of second
cleavage site in US4

<400> 28
Gln Ala Lys Arg Arg
1 5

<210> 29
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: aa of mut.US4
cleavage site

<400> 29
Ala Pro Thr Gln Ala Lys Arg Arg Val Val Gln Arg Glu Lys Ser
1 5 10 15

<210> 30
<211> 1419
<212> DNA
<213> Human immunodeficiency virus

<400> 30

gtagaaaaat	tgtgggtcac	agtctattat	gggggtacctg	tgtggaaaaga	agcaaccacc	60
actctatttt	gtgcatcaga	tgctaaagcc	tatgacacag	aggtacataa	tgctggggcc	120
acacatgcc	gtgtaccacc	agaccctaac	ccacaagaaa	tagtattgga	aaatgtgaca	180
gaaaaatttta	acatgtggaa	aaataaacatg	gtagacagaa	tgcatggagg	tataatcagt	240
ttatgggac	aaagtctaaa	gccatgtgta	aagttaacc	cactctgtgt	tactctacat	300
tgactaat	tgaagaatgc	tactaatacc	aagagttaga	attggaaaaga	gatggacaga	360
ggagaaataa	aaaattgtctc	tttcaagggtc	accacaagaa	taagaaataa	gatgcagaaa	420
gaatctgcac	ttttttataa	actttagtga	gtaccaatag	ataatgataa	tacaagctat	480
aaatgtataa	attgtaacac	ctcagtcatt	acacaggcct	gtccaaaagg	atcctttgaa	540
ccaattccca	tacattattg	tgccccggct	ggttttggca	ttctaaggat	taattgataa	600
aagtctcaatg	gatcaggacc	atgtacaaat	gtcagcacag	tacaatgtac	acattggaatt	660
aggccagtag	tgtaacctca	attgctgtta	aatggcagtc	tagcagaaga	aggggtagta	720
attgatctg	aaaatttccac	agacaatgct	aaaactataa	tagtacagct	gaaggaaatc	780
gtagaaatta	attgtacaag	acctaaacat	aatacaagaa	aaagtataac	tataggaccg	840
gggagagcat	tttatgcaac	aggagacata	ataggagata	taagacaagc	acattgtaac	900
attagtggag	aaaaatggaa	taaacacttta	aaacagtag	ttacaataat	acaagcacaa	960
tttgggaata	aaacaatagt	ctttaagcaa	tctcaggag	gggaccacga	aattgtaaat	1020
cacagtttta	attbtggagg	ggaatttttc	tactgtaatt	caacacagct	ttttaaatag	1080
acttgggaata	atactatagg	gccaaataac	actaatggaa	ctatcacact	cccatgcaac	1140
ataaaacaaa	ttataaacag	gtggcaggaa	gtaggaaaag	caatgtatgc	ccctcccatc	1200
aggagacaaa	ttagatgctc	atcaaatatt	acaggactgc	tattaaacaag	agatgggtgt	1260
aaagagatca	gtaacaccac	cgagatcttc	agacctggag	gtggagatat	gagggaacat	1320
tgaggaagt	aattatataa	atataaagta	gtaaaattgt	agccattagg	agtagcacc	1380
accaagggcaa	agagaagagt	ggtgcagaga	gaaaaaaga			1419

<210> 31

<211> 1932

<212> DNA

<213> Human immunodeficiency virus

<400> 31

gtagaaaaat	tgtgggtcac	agtctattat	gggggtacctg	tgtggaaaaga	agcaaccacc	60
actctatttt	gtgcatcaga	tgctaaagcc	tatgacacag	aggtacataa	tgctggggcc	120
acacatgcc	gtgtaccacc	agaccctaac	ccacaagaaa	tagtattgga	aaatgtgaca	180
gaaaaatttta	acatgtggaa	aaataaacatg	gtagacagaa	tgcatggagg	tataatcagt	240
ttatgggac	aaagtctaaa	gccatgtgta	aagttaacc	cactctgtgt	tactctacat	300
tgactaat	tgaagaatgc	tactaatacc	aagagttaga	attggaaaaga	gatggacaga	360
ggagaaataa	aaaattgtctc	tttcaagggtc	accacaagaa	taagaaataa	gatgcagaaa	420
gaatctgcac	ttttttataa	actttagtga	gtaccaatag	ataatgataa	tacaagctat	480
aaatgtgataa	attgtaacac	ctcagtcatt	acacaggcct	gtccaaaagg	atcctttgaa	540
ccaattccca	tacattattg	tgccccggct	ggttttggca	ttctaaggat	taattgataa	600
aagtctcaatg	gatcaggacc	atgtacaaat	gtcagcacag	tacaatgtac	acattggaatt	660
aggccagtag	tgtaacctca	attgctgtta	aatggcagtc	tagcagaaga	aggggtagta	720
attgatctg	aaaatttccac	agacaatgct	aaaactataa	tagtacagct	gaaggaaatc	780
gtagaaatta	attgtacaag	acctaaacat	aatacaagaa	aaagtataac	tataggaccg	840
gggagagcat	tttatgcaac	aggagacata	ataggagata	taagacaagc	acattgtaac	900
attagtggag	aaaaatggaa	taaacacttta	aaacagtag	ttacaataat	acaagcacaa	960
tttgggaata	aaacaatagt	ctttaagcaa	tctcaggag	gggaccacga	aattgtaaat	1020
cacagtttta	attbtggagg	ggaatttttc	tactgtaatt	caacacagct	ttttaaatag	1080
acttgggaata	atactatagg	gccaaataac	actaatggaa	ctatcacact	cccatgcaac	1140
ataaaacaaa	ttataaacag	gtggcaggaa	gtaggaaaag	caatgtatgc	ccctcccatc	1200
aggagacaaa	ttagatgctc	atcaaatatt	acaggactgc	tattaaacaag	agatgggtgt	1260
aaagagatca	gtaacaccac	cgagatcttc	agacctggag	gtggagatat	gagggaacat	1320
tgaggaagt	aattatataa	atataaagta	gtaaaattgt	agccattagg	agtagcacc	1380
accaagggcaa	agagaagagt	ggtgcagaga	gaaaaaaga	cagtgacgct	aggagctatg	1440
ttccttgggt	tttggggac	agcaggaagc	actatggggc	cacggtcact	gacgctgacg	1500
gtacaggcca	gacaattatt	gtctgtgata	gtgcaacagc	agaacaattt	gtcgagagct	1560
attgaggcgc	aacagcatct	gttgcaactc	acagctctgg	gcatcaagca	gtctcaggca	1620
agagtcctgt	ctgtggaaag	atactcaag	gatcaacagc	tctcagggat	ttgggggtgc	1680


```

ctctggaaaaa  tcatttgcac  cactgctgtg  ccttggaaatg  ctagtgtggag  taataaatctt  1740
ctggatcaga  ttgtgaataa  catgacctgg  atggagtggg  agagagaaat  tgacaattac  1800
acaaacttaa  tatacacttt  aattgaagaa  tcgcagaacc  aacaaagaaa  gaatgaacaa  1860
gaattattag  aattggataa  gtgggcaagt  ttgtggaatt  ggtttgacat  atcaaaatgg  1920
ctgtgtata  ta  1932

```

<210> 32

<211> 2457

<212> DNA

<213> Human immunodeficiency virus

<400> 32

```

gtagaaaaat  tgtgggtcac  agtctattat  ggggtacctg  tgtgaaaga  agcaaccacc  60
actctatttt  gtgcatacga  tgctaaagcc  tatgacacag  aggtacataa  tgtctggggc  120
acacatgcct  gtgtaccacc  agaccttaac  ccacaagaaa  tagtattgga  aaatgtgaca  180
gaaattttta  acatgtggaa  aaataacatg  gtagaacaga  tgcataggga  tataatcagt  240
tattggggtc  aaagtctaaa  gccatgtgta  aagttaaccc  cactctgtgt  tactctacat  300
tgactcaatt  tgaagaatgc  tactaatacc  aagagttagta  attggaaga  gatggacaga  360
ggagaaataa  aaaattgtct  ttccaaggct  accacaagca  taagaataaa  gatgcagaaa  420
gaattgtcac  tttttataaa  acttgatgta  gtaccaatag  ataattgata  taacagctat  480
aaattgtata  attgtaacac  ctacgtcatt  acacaggcct  gtccaagggt  atcctttgaa  540
ccaattccca  tacattattg  tgccccggct  ggttttgcca  ttctaaagtg  taatgataag  600
aagtccaatg  gatcaggacc  atgtacaagt  gtcagcacag  tacaattgat  acatggaatt  660
agggcagtag  tgctcaactc  atgtctgtta  aatggcagtc  tagcagaaga  aggggtagta  720
attagatgtg  aaaatttcc  agacaatgct  aaactataaa  tagtcagct  gaaggatct  780
gtagaataat  attgtacaag  acctaacat  aatacaagaa  aagttataac  tataggaccg  840
gggagagcat  tttatgcaac  aggagacata  ataggagata  taagacaagc  acattgtaac  900
attagtggag  aaaaatggaa  taacacttta  aaacagatag  ttacaaaatt  acagtcacaa  960
tttgggaata  aacaatagat  cttaagcaa  tctcaggagg  ggggcccgaa  aattgtaatg  1020
cacagtttta  attgtggagg  ggaatttttc  tactgtaatt  caacacagct  ttttaatagt  1080
acttggaaat  atactatagg  gccaaataac  actaatggaa  ctatcacact  cccatcgaca  1140
ataaaacaaa  ttataaacag  gtggcaggaa  gtaggaaaag  caatgtatgc  cctccatct  1200
agaggacaaa  ttgatgtctc  atcaaatatt  acaggactgc  tattacaagc  agatgtgtgt  1260
aaagagatca  gttaacaccac  cgagatcttc  agacctgggt  gtggagatat  gagggacatt  1320
tgggaagtg  atatatataa  atataaagta  gtaaaaattg  agccattagg  agtgcaccc  1380
accaaggcaa  agagaagagt  ggtgcagaga  gaaaaaagag  cagtgcagct  agggagattg  1440
ttccttgggt  tcttggggag  agcagggaag  actatggggc  cagcgtcact  gacgtgagct  1500
gtacaggcca  gcaattatt  gtctggtata  gtgcaacagc  agaacaattt  gctgagagct  1560
attgaggcgc  aacagcatct  gtgtgcaact  acagctctgg  gcatcaagca  gctccaggca  1620
agagctctgg  ctgtggaaag  atacctaaag  gatcaacagc  tcttaggatt  ttgggggtgc  1680
tctgaaaaac  tcatttgcac  cactgctgtg  ccttgaaatg  ctagtgtggg  taataaatct  1740
ctgtgacata  ttgtgaataa  catgacctgg  atggagtggg  agagagaaat  tgacaattac  1800
acaaacttaa  tatacacttt  aattgaagaa  tcgcagaacc  aacaaagaaa  gaatgaacaa  1860
gaattattag  atttggataa  gtgggcaagt  ttgtggaatt  ggtttgacat  atcaaaatgg  1920
ctgtgtata  taataaatatt  cataatgata  gtaggaggtt  tagtagtttt  aagtagatt  1980
ttactctgct  ttctatagt  gaattagagt  aggcaggagt  actcaccatt  atcatctcag  2040
accctctcc  cagccccaag  gggaccgcac  aggccgaag  gaatcgaaga  agaagttgga  2100
gagagagaca  gagacagatc  cagtcacatta  gtgcattgat  tattagcact  catctgggac  2160
gatctacgag  gctctgtcct  ctccagctac  caccgcttga  gagacttaac  ctgtatgca  2220
gcgaggattg  tggaaattct  gggacgcagg  ggggtgggaag  cctccaagta  ttgggggaat  2280
ctcctgcagt  attgatttca  ggaactaaag  aatagtcgtg  ttagtgtgtt  tgatgccata  2340
gctatagcag  tagctgaggg  gacagatagg  attatagaag  tagcacaagc  aattgtgata  2400
gcttttctcc  acatacctag  aagaataaga  cagggtcttg  aaagggtctt  gctataa  2457

```

<210> 33

<211> 1453

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp120.modSF162

<400> 33

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagttcttc ttctgccag cgccgtggag aagctgtggg tgaccgtgta ctacggcggt 120
ccgtgtgtgga aggaaggccac caccacctgt ttctgcgcca gcgacgccaa ggctctacgac 180
accgaggtgac acaacgtgtg ggccaccacc gctgtgctgc ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgtgag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gctgacacct gcactgcacc aacctgaaga acgccaccaa accaaacggc 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtac gccctgttct acaagctgga cgtgaaacggc 540
atcgacaaac acaacaccag ctacaagctg atcaactgca acaccagcgt gatcaccagg 600
gctgcgccca aggtgagctt cgagcccatc cccatccaat actgcgcccc cgccgcttcc 660
gccatcctga agtgacaaga caagaagttc aacggcagcg gccctgtcac caacgtgagc 720
accgtgcagt gcaaccacgg catccgcccc gtggtagaca cccagctgct gctgaaacggc 780
agcctggcgg aggaggggct ggtgatccgc agcgagaact tcaccgacaa cgccaagacc 840
atcatcgtgc agctgaagga gacgtggag atcaactgca cccgccccaa caacaacacc 900
cgcaagaagca tcaccatcgg ccccgccgcg gcccttctacg ccaccggcga catcatcggc 960
gacatccgca acaagatgca caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtagaca agctgcagcg ccagttcggc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcggcgacc ccgagatcgt gatgcacagc ttcaactgcg gcggcgagtt ctcttactgc 1140
agcaccatca cctgcacctg ccgcatcaag cagatcatca accgtgggca ggaagtgagg 1200
aaggccatgc acgccccccc catccggggc cagatccgct gcagagcaca catcacggcg 1320
ctgctgctga ccgcgcagcg ggcgaaggag atcagcaaca ccaccgagat ctctcgcccc 1380
ggcggcgcg acatgcgcga caactggcgc agcgagctgt acaagtagaa ggtggtgaag 1440
atcgagcccc tgg                                     1453

```

<210> 34

<211> 1387

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp120.modSF162.delV2

<400> 34

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagttcttc ttctgccag cgccgtggag aagctgtggg tgaccgtgta ctacggcggt 120
ccgtgtgtgga aggaaggccac caccacctgt ttctgcgcca gcgacgccaa ggctctacgac 180
accgaggtgac acaacgtgtg ggccaccacc gctgtgctgc ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgtgag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gctgacacct gcactgcacc aacctgaaga acgccaccaa accaaagagc 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtggtggcg 480
ggcaagctga tcaactgcaa caccagcgtg atcaccagcg cctgccccaa ggtgagcttc 540
gagcccatcc ccatccacta ctgcgcccc gcggttctcg ccatcctgaa gtgcaacgac 600
aagaagttca acggcagcgg ccctgcacc aacgtgagca ccgtgcagt cccccacggc 660
cttcgccccg tggtagacac ccagctgctg ctgaacggca gctgtggcga ggaagggcgt 720
gtgatccgca gcgagaactt caccgacaac gccaaagaca tcactgtgca cgtgaaaggag 780
agcgtggaga tcaactgcac ccgccccaac aacaacacc gcaagagcat caccatcgcc 840
cccgcccgcg ccttctacgc caccggcgac atcatcgggc acatccgca ggcacactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgaccaa gctcgaggcc 960
cagttcgcca acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcaccca gctgttcaac 1080
agcactcgga acaacacctt cggccccaac aacaccaacg gcaacctcac cctgcctcgc 1140

```

```

cgcacatcaagc agatcatcaa ccgctggcag gaggctgggca aggccatgta cgccecccc 1200
atccgcggccc agatccgctg cagcagcaac atcaccggcc tgcctgtgac ccgcgacggc 1260
ggcaaggagga tcagcaaacac caccgagatc ttccgccccg gcggcgccga catgcgcgac 1320
aactggcgca gcgagctgta caagtacaag gtggtgaaga tcgagccctt gggcgctggcg 1380
ccccacca
1387

```

<210> 35
 <211> 1323
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: gp120.modSF162.delV1V2

```

<400> 35
gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cgagttcttc ttctgccacg cgcctgggag aagctgtggg tgaccgtgta ctacggcgcg 120
cccggtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgccaa ggccatcgac 180
accgaggtgc acaacgtgtg ggccaaccac gctctgctgc ccaccgaccc caaccgccac 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggcacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtggggcg cggcaactgc cagaccagcg tgatcaccca ggccctgcce 420
aaggtgagct tcgagcccat ccccatccac tactgcgccc ccgcggctt gcccatcctg 480
aagtgcacac acaagaagtt caacggcagc ggccccctga ccaacgtgag caccgtgcag 540
tgaccacacg gcattccgccc cgtggtgagc acccagctgc tgcgaaagg cagcctggcc 600
gaggagggcg tggatgatcc cagcgagaga ttaccgcaca acgccaagac catcatcgtg 660
cagctgaagg agagcgtgga gatcaactgc acccgcccca acaacaacac ccgcaagagc 720
atcaccatcg gcccgcggcg cgcctttctc gccaccggcg acatcatcgg cgacatccgc 780
caggcccact gcaacatcag cggcgagaag tggaaacaaca ccctgaagca gatcgtgacc 840
aagctgcagg ccagtttcgg caacaagacc atcgtgttca agcagagcag cggcgggcag 900
cccgagatcg tgatgcacag ctccaactgc ggccgagagt tcttctact ccaacgaccc 960
cagctgttca acagcaccct gaacaacacc atcgcccca acaaccacca cggcaccatc 1020
accctgcctt ccgcacatca gcagatcacc aaccgctggc aggaggtggg caaggccatg 1080
tacgcccccc ccattccggc ccagatccgc tgcagcaga acatcacccg cctgctgctg 1140
accgcgacg ccggcaagga gatcagcaac accaccgaga tcttccgccc gggcgggcgg 1200
gacatgcgcg acaactggcg cagcgagctg tacaagtaca agtggtgtaa gatcgagccc 1260
ctggcgctgg cccccacca ggccaagcgc cgcgtgtgct agcgcgagaa ggcgttaact 1320
gag
1323

```

<210> 36
 <211> 2025
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: gp140.modSF162

```

<400> 36
gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cgagttcttc ttctgccacg cgcctgggag aagctgtggg tgaccgtgta ctacggcgcg 120
cccggtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgccaa ggccatcgac 180
accgaggtgc acaacgtgtg ggccaaccac gctctgctgc ccaccgaccc caaccgccac 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggcacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt cgtgacacct gcactgcacc aacctgaaga acgccaacca cccaagagc 420
agcaactgga aggaagatgga ccgcggcgag atcaagaact cgagcttcaa ggtgaccacc 480
agcattccgca acaagatgca gaaggagtag gccctgttct acaagctgga cgtgtgtccc 540
atcgacaacg acaaccacag ctacaagctg atcaactgca acaccagcgt gatcaccagg 600

```

```

gcttgcceca aggtgagctt cgagcccatc cccatccact actgcgcccc cgcggcttc 660
gccatcttga agtgcaacga caagaagttc aacggcgagc gccctgcac cactgtgagc 720
accgtgcagt gcaaccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc 780
agcctggccg agggaggcgt ggtgatccg agcgagaact tcaccgacaa cgccaagacc 840
atcatctgtc agctgaagga gagcgtggag atcaactgca cccgccccaa caacaacacc 900
cgcaagagca tcaecatcgg ccccgccgcg gcccttctacg ccaccggcga catcatcggc 960
gaatctcgcc agggccactg caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtgacca agctcgagcg ccaagtccgc aacaagacca tcgtgttcaa gcgagcgagc 1080
ggcgcgacc ggagatcgt gatgcacagc ttcaactgct ttctactgc 1140
aacagacccc agctgttcaa cagcacctgg aacaacacca tcggccccaa caaccacnac 1200
ggccacatac cctgcccctg ccgcataca cagatcatca accgctggca ggaggtggcg 1260
aaggccatgt acgccccccc catcccgccg cagatccgct gcagcagcaa catcacccgc 1320
ctgctgctga cccgcgagcg cggcaaggag atcagcaaca ccaccgagat ctctccgccc 1380
ggcgccggcg acatcgcgga caactggcgc agcgagctgt acaagtacaa ggtggtgaag 1440
atcgagcccc tggcgctggc ccccaaccaag gccaaagccc tcgtggtgca gcgcgagaag 1500
cgcgccgtga cctggggcgc catgttctct ggcttctctg gcgcgcccgg cagcacccatg 1560
ggcgcccgca gctgaccctt gaccgtgcag gcccgccgag tgcgtgagcg catcgtggcg 1620
cagcagaaca acctgctcgc gccatcgag gccccagcgc acctgctgca gctgaccctg 1680
tggggcatca agcagctgca gggcccgctg ctggccctgg agcgctacct gatgaccag 1740
cagctgctgg gcatctgggg ctgacggcgc aagctgatct gcaccaaccg cgtgccctgg 1800
aacgccagct ggagcaacaa gagcctggag cagatctgga acaacatgac cttgattggg 1860
tgggagcgcg agatcgacaa ctacaccaac ctgatctaca cctgatcgca ggagagccag 1920
aacacgacg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
aactggttcg acatcagcaa gtggtgtgg tactctaac tcgag 2025

```

<210> 37
 <211> 1944
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp140.modSF162.delV2

```

<400> 37
gaattcgcca ccatggatgc aatgaagaga gggtctgct gtgtgctgt gctgtgtgga 60
cgagtcttcc ttctgcccag cgcgtggag aagctgtggg tgacctgtga ctacggcgtg 120
cccgtgtgga aggaggccac caccacctg ttctgcgcca gcgacgcaaa ggcttacgac 180
accgaggtgc acaactgtgt gggcaaccac gcctcgctgc ggaagaacaa catgtgtggag 240
gagatcgtag tggagaacgt gaccgagaa ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagcgt 360
acccccctgt gctgtagcct gcactgcacc aaactgaaga acgcacacaa caccagagc 420
agcaactgga aggagatgga cgcgcgagag atcaagaact gcagcttcaa ggtggcgccc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccaggg cctgccccaa ggtgagcttc 540
gagccccatc ccatccaacta ctgcgccccg gccggtcttc ccatcctgaa gtgcaacagg 600
aagaagtgtc acggcgagcg cccctgcacc aacgtgagca ccgtgcagtg caccacacgc 660
atccgccccg tggtaggcac ccagctgctg ctgaacggca gcctggccga ggagggcgtg 720
tgatcccgca gcgagaactt caccgacaac gccaaagcca tcatcgtgca gctgaagagg 780
agcgtggaga tcaactgcac ccgccccaac aacaacacc ccgaagacat caccatcggc 840
cccgcccgcg ccttctaocg caccggcgag atcatcgccg acatcccgca ggcccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgaccaa gctgcaggcg 960
cagttgcgca acaagaccat cgtgttcaa cagagcagcg gcgcgacccc cagatcgtg 1020
atgcacagct tcaactcgcg cggcgagttc ttctactgca acagacacca gctgttcaac 1080
agcacttgga acaacaccat cggccccaac aacaccaacg gcaccatcac cctgcctcgc 1140
gcatacaagc agatcatcaa ccgctggcag gagggtggga agggcatgta cgcgccccc 1200
atcccgccgc agatcccgct cagcagcaac atcaccgccc tcgtgctgac ccgacacgc 1260
agccaggaga tcagcaaac caccgagatc ttccgccccg gcggcgcgga catcgcgcac 1320
aactggcgca gcgagctgta caagtacaag gtgggtgaag tcgagccccct gggcgtggcc 1380
cccaccaagg ccaagcgccg cgtggtgcag cgcgagaagc gcgcgctgac cctggcgccc 1440

```

atgttctctgg	gcttctctggg	cgccgcccggc	agcaccatgg	gcgcccgcag	cctgaccctg	1500
accgtgcagg	cccgcacgct	gctgagcggc	atcgctgcagc	agcagaacaa	cctgtctgctg	1560
gccatgcagg	cccagcagca	cctgtctgcag	ctgaccgtgt	ggggcatcaa	gcagctgcag	1620
gcccgcgtgc	tgccgcgtgga	gcgctacctg	aaggaccagc	agctgctggg	catctctgggg	1680
tgacgcggca	agctgatctg	caccacgcgc	gtgcctctgga	acgccagctg	gagcaacaag	1740
agcctggacc	agatctggaa	caacatgacc	tggaaggagg	gggagcgcca	gatcgacaac	1800
tacaccaacc	tgatctacac	cctgatcgag	gagagccaga	accagcagga	gaagaacagag	1860
caggagctgc	tgagagctgga	caagtggggc	agcctgtgga	actggttcga	catcagcaag	1920
tggtgtgtgt	acatctaact	cgag				1944

<210> 38

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.modSF162.delV1/V2

<400> 38

gaatttcgcca	ccatggatgc	aatgaagaga	gggctctgct	gtgtgctgct	gctgtgtgga	60
gcagttcttcg	tttgcgccag	cgccgtggag	aagctgtggg	tgaccgtgta	ctacggcgctg	120
cccgtgtgga	aggaggccac	caccaccctg	ttctgcgcca	gcgacgcgca	ggcctacagc	180
accgaggtgc	acaacgtgtg	ggccacccac	gctgtcgctg	ccaccgaccc	caaccgccag	240
ggagatcgtgc	tggaagaact	gaccgagagc	ttcaacatgt	ggaagaacaa	catgtgtggag	300
cagatgcacg	aggacatcat	cagcctgtgg	gaccagagcc	tgaagccctg	cgtgaaagctg	360
acccccctgt	gcgtgaccct	gcactgcacc	aacctgaaga	acgccaccac	caccaagagc	420
agcaactgga	aggagatgga	cgccggcgag	atcaagaact	gcagcttcaa	ggtgggcggcc	480
ggcaagctga	tcaactgcaa	caccagcgtg	atcaccacag	cctgcgccaa	ggtgagcttc	540
gagcccatcc	ccatccacta	ctgcgcgcc	gcggcttcg	ccatcctgaa	gtgcaacgac	600
aagaagttcaa	acggcagcgg	cccctgcacc	aacgtgagca	ccgtgcagtg	cacccacggc	660
atccgcgcgc	tggtgagcac	ccagctcgtg	ctgaacggca	gcttgccgga	ggaggggcgtg	720
gtgatccgca	gcgagaactt	caccgacaac	gccaaagaca	tcatcgtgca	gctgaaggag	780
agcgtggaga	tcaactgcac	cgcccccaac	aacaacaccc	gcaagagcat	caccatcggc	840
cccggccgcg	ccttctaacc	caccggcgac	atcctcgccg	acatccgcca	ggcccaactgc	900
aacatcagcg	gcgagaagtg	gaacaacacc	ctgaagcaga	tcgtgaccaa	gctgcagggc	960
cagtttcgcca	acaagaccat	cgtgttcaag	cagagcagcg	gcggcgaccc	cgagatcgtg	1020
atgcacagct	tcaactgcgg	cggcgagttc	ttctactgca	acagcaccca	gctgttcaac	1080
agcaactcga	acaacaccat	cggccccaac	aacaccaacg	gcaccatcac	gctgcctctg	1140
cgcatacaag	agatcatcaa	cgcgtggcag	gaggtggggca	aggccatgta	cgcccccccc	1200
atccggcgcc	agatcccgctg	cagcagcaac	atcacccggc	tgctgtgcag	ccgcgacggc	1260
ggcaaggaga	tcagcaaacac	caccgagatc	ttccgccccg	gcggcgccga	catgcgcgac	1320
aactcggcca	cgagactgta	caagtacaag	gtgggtgaag	tcgagccctg	ggcgctggcc	1380
cccaccaagg	ccaagccggc	cgtggtgcag	cgcgagaagc	gcgcccgcag	cctgggcggc	1440
atgttctctg	gcttctctggg	cgccgcggcg	agcaccaatg	gcgcccgcag	cctgacctctg	1500
accgtgcagg	cccgcacgct	gctgagcggc	atcgctgcagc	agcagaacaa	cctgctgcgc	1560
gccatcgagg	cccagcagca	cctgctgcag	ctgacccgtg	ggggcatcaa	gcagctgcag	1620
gcccgcgtgc	tgccgtgga	gcgctacctg	aaggaccagc	agctcgtggg	catctggggc	1680
tgacgcggca	agctgatctg	caccacgcgc	gtgcctctgga	acgccagctg	gagcaacaag	1740
agcctggacc	agatctggaa	caacatgacc	tggaaggagg	gggagcgcca	gatcgacaac	1800
tacaccaacc	tgatctacac	cctgatcgag	gagagccaga	accagcagga	gaagaacagag	1860
caggagctgc	tgagagctgga	caagtggggc	agcctgtgga	actggttcga	catcagcaag	1920
tggtgtgtgt	acatctaact	cgag				1944

<210> 39

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.mut.modSF162

<400> 39

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagttcttc ttctgcccag cgcctggag aagctgtggg tgaccgtgta ctaccggctg 120
cccgcttgga aggaaggccac caccaccctg ttctgcccga gcgacgcgaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac cctctgctgc ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggaacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gctgacccct gcactgcacc aacctgaaga acgcccacaa caccagaagc 420
agcaactgga aggaatgga ccgcgccgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatcgcca acaagatgca gaaggagtag cccctgttct acaagctgga cgtggtgccc 540
atcgacaacg acaaccacag ctacaagctg atcaactgca acaccagctg gatcacccag 600
gctgtcccga aggtgagctt cgagcccact cccatccact actgcccctc cgcgggtctc 660
gccatcttga agtgcaacga caagaagttc aacggcagcg gccctgcac caactgagcg 720
accgtgcagt gacccacacg catccgcccc gtggtgagca cccagctgct gctgaacggc 780
agcctggccg aggaaggcgt ggtgatccgc agcgagaact tcaccgacaa gcacgaagac 840
atcatctgtc agctgaagga gagcgtggag atcaactgca cccgcccacaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgccgcg gcccttctac caaccgagca catcatcggc 960
gacatccgct aggcctcactg caacatccag ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtgacca agcgtcaggg ccagttccgc aacaagacca tctgtgtcaa gcagagcagc 1080
ggcgccgacc ccgagatcgt gatgcacagc ttcaactgcy ggcgcgagtt ctctactgc 1140
aaccgacccc agctgttcaa cagcacctgg aacaacacca tcggcccacaa caacaacacc 1200
ggcacctaca cccgtcccct gcgcatacag cagatcatca accgttgcca ggaggtgggc 1260
aaggccatgt accgcccccc catccggcgc cagatccgct gcagcagcaa catcacccgc 1320
ctgctcttga cccgcgacgg cggcaaggag atcagcaaca ccaccagat ctccgcccc 1380
ggcgccggcg acatcgccga caactggcgc agcgagctgt acaagtacaa ggtggtgaag 1440
atcgaccccc tggcgctggc ccccaaccaag gccaaagcgc cgtggtgtga gcgcgagaag 1500
agcgccctga cctggggcgc catgttctct ggttctctgg gcgcgcggcg cagcaccatg 1560
ggcgcccgca gccctgacct gaccgtgcag gccgcgcagc tctgagcgg catcgtcag 1620
cagcagaaca acctgctgcy cgcctcagag gccccagcgc acctgtgtga cgtgaccgtg 1680
tggggcatca agcagctgca ggcccgcgtg ctggccgtgg agcgtacct gaagagaccag 1740
cagctgtctg gcactctggg ctgcagcggc aagctgatct gcaccacgc cgtgccctgt 1800
aacgcagcgt ggaagcaaca gagcctggac cagatctgga acaacatgac ctggatggag 1860
tgggagcgcg agatcgacaa ctacaccaac ctgactcata cccctgatga ggagagccag 1920
aacccagcag agaagaacga gcaggagctg ctggagctgt acaagtgggc cagcgtgtgg 1980
aactggttgc acatcagcaa gtggctgtgg tacatctaac tcgag 2025

```

<210> 40

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.mut.modSF162.delV2

<400> 40

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagttcttc ttctgcccag cgcctggag aagctgtggg tgaccgtgta ctaccggctg 120
cccgcttgga aggaaggccac caccaccctg ttctgcccga gcgacgcgaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac cctctgctgc ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggaacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gctgacccct gcactgcacc aacctgaaga acgcccacaa caccagaagc 420
agcaactgga aggaatgga ccgcgccgag atcaagaact gcagcttcaa ggtgagccgc 480
ggcaagctga tcaactgcaa caccagcgtg ataccaccag cctgcgccaa ggtgagcttc 540

```

```

gagcccatcc ccatccacta ctgcgcccc ggcggtctcg ccatcctgaa gtgcaacgac 600
aagaagtcca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacacggc 660
atccgccccg tggtagacac ccagctgctg ctgaacggga gccctggcga ggaggcgctg 720
gtgatccgca ggagaaactt caccgacaac gccaaagcca tcactgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccgccccaac aacaacaccc gcaagagcat caccatcggc 840
cccgcccgcg ctttctacgc caccggcgac atcatcgcg acatccgcga ggcccaactgc 900
aacatcagcg agagaagtg gaacaacacc ctgaagcaga tcgtgaccaa gctgcagcgg 960
cagtttcggca acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactcgcg cgcgcgagttc ttctactgca acagaccca cgtgttcaac 1080
agcaccctgga acaacacat cggccccaac aacacacacg gcacatcac cctgcctcgc 1140
cgcatcaagc agatcataaa ccgctggcag gaggtgggca aggccatgta cggccccccc 1200
atcccgcgcc agatccgctg cagcagcaac atcaccggcc tgctgctgac ccgcagacggc 1260
ggcaagggga tcagcaacac caccgagatc ttccgccccg cgggcgcgca catgcgcgac 1320
aactgcgcga gcgagctgta caagtacaag tgggtgaaga tcgagccccct gggcggtggcc 1380
cccaccaagg ccaagcgccg cgtggtgcag cgcgagaaga gcgccgtgac cctggggcgc 1440
atgttctctg gcttctctgg cgcggcgccg agcaccatgg gcgcccgac cctgaccctg 1500
accgtgcagg ccgcacagct gctgagcgcc atcgtgcagc agcagaacaa cctgtgtcgc 1560
gccatcgagg ccacgacgca cctgctgcag ctgacccgtg ggggcatcaa cgtgctgcag 1620
cccgcgctgc tggcgctgga gcgctacctg aaggaccagc agctgctggg catctggggc 1680
tgcagcgcca agctgactcg caccaccgcc gtgccctgga acgacagctg gtcgacaag 1740
agcctggacc agatctggaa caacatgacc ttggtggagt gggagcgcca gatcgacaac 1800
tacaccaacc tgatctacac cctgatcgag gagagccaga gaagaacagg 1860
caggagctgc ttgagctgga caagtggggc agcctgtgga actggttcga catcagcaag 1920
tggctgtggt acatctaact cgag

```

<210> 41

<211> 1836

<212> DNA

<213> Artificial Sequence

<220>

<221> Description of Artificial Sequence:
 gp140.mut.modSF162.delV1/V2

<400> 41

```

gaattcgcca ccatggatgc aatgaagaga gggtctgtgt gctgtgctgt gctgtgtgga 60
gcagctcttcg ttctgcccag cgcggtggag aagctgtggg tgaccgtgta ctacggcgctg 120
cccgcttgga aggaggccac caccaccctg ttctgcgcca gcgacgcaa ggccacagac 180
accgaggtgc acaactgtgt ggccacccac gcctgcgtgc ccacgcgccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgaacg agagcatcat cagcctgtgg gaccagagcg tgaagccctg cgtgaagctg 360
acccccctgt gcgtggggcg cggcaactgc cagaccagcg tgatcaccca ggccctcccc 420
aagtgagctc tgcagcccat ccccatccac tactgcgccc ccgcgggctt cgcactctgt 480
aagtgcacag acaagaagt caacggcgag ggcctctgca ccaacgtgag caccgtgcag 540
tgacccacag gcctccgccc cgtggtgagc acccagctgc tgctgaacgg cactcgtgcc 600
gaggaggggg tgggtgatcc cagcgagaac ttaccgcgca acgccaagac catcatcgtg 660
cagctgaagc agagcgtgga gatcaactgc acccgcccca caacaacac ccgcaagcag 720
atcaccatcg gccccggcgg cgccttctac gccaccggcg acatcatcgg cgacatcccg 780
caggcccact gcaacatcag cggcgagaag tggaaacaaa cctgaagcga gatcgtgacc 840
aagctgcagg ccaggttcgg caacaagacc atcgtgttca agcagaagcag gggcgggcag 900
cccgagatgc ttgtgcacag ctccaactgc ggcggcgagt ttcttactgc caacagcacc 960
cagctgttca acagcacctg gaacaacacc atcgggccca acaacaccaa cggcaaccatc 1020
acctgcacct gccgcataaa gcagatcatc aaccgtggcg aggaggtggg caaggccatg 1080
tacgcccccc ccacatccggg ccagatcccg tgcagcagca acatcacgg cctgctgctg 1140
accgcgcagc gcggcaagga gatcagcaac accaccgaga tcttcctccc cggcgggcgg 1200
gactatgcgg acaactggcg cagcagagctg tacaagtaca agtggtgtaa gatcgagccc 1260
ctggcgctgc ccccaccaa ggccaagcgc cgcgtggtgc agcgcagaga gagcgcgctg 1320
acctggggcg ccatgttctt gggcttctct ggcgcggcgg cagcacccat gggcgacggc 1380
agcctgaccc tgacgctgca ggcgcggcag ctgctgagcg gcatcgtgca gcagcagaac 1440

```

```

aactctgtctgc gcgccaatcga ggcccagcag cacctgtctgc agctgaccgt gtggggcctc 1500
aagcagctctgc agggccgcgt gctggccgtg gagcgtctacc tgaaggacca gcagctctctg 1560
ggcatcttggg gctgcagcgg caagctgctc tgcaccaccg ccgtgcccctg gaaccgccagc 1620
tggagcaaca agagcctcga ccagatctgg aacaacatga cctggatgga gtgggagcgc 1680
gagatcgaca actacaccaa cctgatctac accctgatcg agggagagcca gaaccagcag 1740
gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactgtgtc 1800
gacatcagca agtggtctgt gtacatctaa ctgcag

```

<210> 42
 <211> 2025
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp140.mut7.modSF162

```

<400> 42
gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtctgt gctgtgtgga 60
cgagctcttcg ttctgccagc cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
ccgtgttggga aggaggccac caccaccctg ttctgcgcca gcgacgcgaa ggcctacgag 180
accgaggttgc acaactgtgt ggccaccacc gctctgctgc ccaccgaccc caaccgccag 240
gagatctgtc tggagaacct gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg agggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt cgtgtacct gcactgcacc aacctgaaga acgcccacaa caccagaagc 420
agcaactcga aggatgtgga ccgcgccgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatcgcga acaagatgca gaaggagtac gcctctgtct acaagctgga cgtggtgtcc 540
atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt galcaccagc 600
gctctcccca agtgtagctt cgagcccatc cccatccact actgcgcccc cgccggtctc 660
gccatcctga agtgcaacga caagaagttc aacggcagcg gccctgtcac caacgtgagc 720
accgtgcagt gcaccaccgg catccgcccc gtggtgagca cccagctgct gctgaacggc 780
agcttgcgcc aggagggcgt ggtgatccg agcgagaact tcaccgacaa cgcaaacgac 840
atcatctgtc agctgaagga gtagctggag atcaactgca cccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgccgcg gctctctacg ccaccggcga catcatcgcc 960
gacatccgcc agggccactg caacatcagc ggcgagaagt ggaacaacac cctgaacgag 1020
atcgtgacca agctcgagcg ccagttcggc aacaagacca tctgtgtcaa gcagagcagc 1080
ggcgcgacc ccgagatcgt gatgcacagc ttcaactgcg gcggcgagtt ttcttactgc 1140
aacagcacc ccagctgtcaa cagcacctgg aacaacacca tcggccccaa caacaccaac 1200
ggcaccatca cctgcacctg ccgcatcaag cagatcatca accgctggca ggaggtggcg 1260
aaggccatgt acgccccccc catccggcgg cagatccgct gcagcagcaa catcaccggc 1320
ctgtctctga ccgcgcagcg cggcaaggag atcagcaaca ccacgagat ctctccccc 1380
ggcgcgccgg acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtgaa 1440
atcgagcccc tgggcgtggc cccaccacaa gccatcagca cgctggtgca gagcgagaag 1500
agcgcccgta ccttggcgcg catgttctct ggtctctgtg gcgcgcccgg cagcaccatg 1560
ggcgcccgca gctgacctt gaacctgcag gcccgccagc tgctgagcgg catcgtgcag 1620
cagcagaaca acctgtctgc cgccatcgag gcccgagcgc acctgtctga cctgaacctg 1680
tggggcagca agcagctgca ggcggcgctg ctggcgctgg agcctacct gaaggaccag 1740
cagctctgct gcatctgggg ctgcagcgcc aagctgatct gcaccaccgc cgtgcctcgg 1800
aacgccagct agcagcaaca gagcctggac cagatctgga acaacatgac ctggatggag 1860
tgggagtcgc agatcgacaa ctacaccaac ctgatctaca cctgatgca ggagagccag 1920
aacagcagc agaaagacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
aactggttgc acatcagcaa gtggctgtgt tactctaac tcgag

```

<210> 43
 <211> 1944
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut7.modSF162.delV2

<400> 43

```

gaattcgccca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
gcagctcttcg ttctgcccgag cgcctgtggag aagctgtggg tgaccgtgta ctacggcggtg 120
cccggtgtgga aggaggccac caccacctgt ttctgcgcca gcgacgccaa ggctctacgac 180
accgaggtgcg acaacgtgtg gggccaccac gcctgcgtgc ccaccgaccc caaccgccag 240
gagatcgtgcg tggagaacgtg gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcgacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accgccctgtg gcgtgacctt gcactgcacc aacctgaaga acgccacca caccgaagcg 420
agcaactgga aggagatgga ccgcgggcgag atcaagaact gcagcttcaa ggtggcgccc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccaggg cctgccccta ggtgagcttc 540
gagcccatcc ccatccta ctgcgcccc gcccgtctcg ccatcctgaa gtgcaacgac 600
aagaagtcca accgagcggg cccctgcacc aacgtgagca ccgtgcagtg caccacggc 660
atccgccccg ttggtgagcac ccagctgctg ctgaacggca gcctggcgga ggaggcgctg 720
gtgatccgca gcgagaactt caccgacaac gccaaagca tcatcgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccgccccaac aacaacaccc gcaagagcat caccatcggc 840
ccgcgccggcg cctctacgc caccggcgac atcatcgcca acatccgca ggcccactgc 900
aaatcatcggc gcgagaagtg gaacaacacc ctgaagcaga tctgtacca gctgcaggcc 960
cagttcgccga acagacccat cgtgttcaag cagagcagcg cggcgagccc cagagctcgt 1020
atgcacagct tcaactgcgg cggcgagttc ttctactga acagaccca gctgttcaac 1080
agcacctgga acacacccat cggccccaac aacaccaacg gccacatcac cctgcccctg 1140
cgcatccaga agatcatcaa ccgctggcag gaggtggga aggccatgta cgcccccccc 1200
atcccgccggc agatccgctg cagcagcaac atcaccggcg tgcgtgctgac ccgcgacggc 1260
ggcaaggaga tcagacaacac caccgagatc ttccgccccg cggcgcgcca catgcgcgac 1320
aactgcgcga gcgagctgta caagtacaag gtggtgaaga tcgagccccc gggcgtggcc 1380
cccacaagg ccagcatcag cgtggtcgag agcgagaaga gcgcccgtgac cctggggcgcc 1440
atgttctctgg gcttctctgg cgcgcggcgc agcaccatgg gcgcccgcag cctgaccctg 1500
accgtgcagg ccgccagact gctgagcggc atcgtgcagc agcagaacaa cctgtctgac 1560
gccatcgagg ccacgcagca cctgctgcag ctgaccgtgt ggggcatcaa gcatgctgag 1620
gcccgctgic tggccgtgga gcgctacctg aaggaccagc agctgctggg catctggggc 1680
tgacgcgcga agctgatctc caccaccgcg gtgccctgga accgcagctg gagcaacaag 1740
agcctggacc agatctgaa caacatgacc tggatggagt gggagcgcca gatcgacaa 1800
tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtgggca agcctgtgga actggttoga catcagcaag 1920
tggctgtggt acatctaact cgag

```

<210> 44

<211> 1836

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut7.modSF162.delV1/V2

<400> 44

```

gaattcgccca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
gcagctcttcg ttctgcccgag cgcctgtggag aagctgtggg tgaccgtgta ctacggcggtg 120
cccggtgtgga aggaggccac caccacctgt ttctgcgcca gcgacgccaa ggctctacgac 180
accgaggtgcg acaacgtgtg gggccaccac gcctgcgtgc ccaccgaccc caaccgccag 240
gagatcgtgcg tggagaacgtg gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcgacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accgccctgtg gcgtggggcg cggcaactgc cagaccagcg tgatcaacca ggccctgccc 420
aagtgagctg tcgagcccat ccccatccac taactgcgcc cgcggcgctt cgccatcctg 480
aagtgcaacg caagaagtt caacggcagc ggcccctgca ccaacgtgag cagcgtgcag 540
tgacccacac gcatccgccc cgtggtgagc acccagctgc tgcgtgaacg accgtggcc 600
gaggagggcg tgggtgatcg cagcagaaac ttaccggaca acgccaagac catcatcgtg 660

```

```

cagctgaagg agagcgtgga gatcaactgc acccgcccca acaacaacac ccgcaagagc 720
atcaccatcg gccccggccg cgccctctac gccaccggcg acatcatcgg cgacatccgc 780
caggccccact caaacatcag cggcgagaag tggaaacaac cctcgaaaga gatcgtgacc 840
aagctgcagg ccaggttcgg caacaagacc atcgtgttca agcagagcag cggcggcgagc 900
cccgagatcg tgatgcacag ctccaactgc ggccggcgagt tcttctactg caacagcacc 960
cagctgttca acagcaccct gaacaacacc atccggccca acaaccacaa cggcaccatc 1020
accctgcctt gccgcatcaa gcagatcacc aaccgctggc aggaggtggg caaggccatc 1080
tacgcccccc ccaltccggg ccagatcccg tcagcagca acatcacccg cctgctgctg 1140
accgcgagcg ggggcaagga gatcagcaac accaccgaga tcttcggccc cggcggcgcc 1200
gacatcgccg caactcgccg cagcgagctg tacaagtaca aggtggtgaa gatcgagccc 1260
ctggggtcgg ccccccacaa ggccatcagc agcgtggtgc agagcgagaa agcgcccgctg 1320
accctggggg ccatgttctt gggcttctgt ggccggcccg gcagcaccat gggcgcccg 1380
agcctgaccc tgaccgtgca ggcccgccag ctgctgagcg gcactgctga cagcagaaac 1440
aacctgtcgc gcgccatcga ggcccgccag cactgctgc agctgacgct gtggggcatc 1500
aagcagctgc agggccgctg cctggccgtg gagcgtacc tgaaggacca cagctgctg 1560
ggcatctcgg gctgcagcgg caagctgctg tgcaccaccg cgtgcccctg gaaccggcag 1620
tgagagaaca agagcctgga ccagatctgg aacaacatcg ctggatgga gtgggagcgc 1680
gagatcgaca actacaccaa cctgatctac accctgatcg aggagagcca gaaccagcag 1740
gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctggt gaactgtgtc 1800
gacatcagca agtggctgtg gtacatctaa ctccag 1836

```

<210> 45

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.mut8.modSF162

<400> 45

```

gaatttcgcca ccatggatgc aatgaagaga gggtctgtgt gtgtgctgct gctgtgtgga 60
gcagttcttcg ttctgcccag cgccgtggag aagctgttgg tgaccgtgta ctacggcggt 120
cccgtgtgga aggaggccac caccaccctg ttctgcccga gcgacgcaa ggccctacgac 180
accgaggtgc acaacgtgtg ggccaccacc gctcgctgct ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaa ttcaaatgt ggaagaacaa catggtggag 300
cagatgcacg aggcacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt cgctgacctt gcactgcacc aacctgaaga acgcaccaca cccaagagctg 420
agcaactgga agagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtag gccctgttct acaagctgga cgtggtgcc 540
atcgacaacg acaaacaccag ctacaagctg ctcaactgca acaccagcgt gatcaccagc 600
gcccctccca aggtgagctt cgagcccatc cccatccact actgcgcccc cgccggtctc 660
gctctcctga agtgcaacga caagaagtgc aacggcagcg gccctgtgac caacgtgagc 720
accgtgcagt gcacccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc 780
agcctggccg aggaggcggt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc 840
atcatcgtgc agctgaagga gagcgtggag atcaactgca cccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgccgcg gccttctacg ccacgggga catcatccgc 960
gacatccgcg aggcaccact caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtgacca agctgcaggc ccagttcgcc aacaagacca tegtgttcaa cgagagcagc 1080
ggcgcgaccg ccgagatcgt gatgcacagc ttcaactcgc cggcgagagt cttctactgc 1140
aacagcaccg agctgttcaa cagcacctgg aacaacacca tcggccccaa caacacaaac 1200
ggcaccatca cctgcgccct ccgcatcaag cagatcatca accgctggca ggaggtggcg 1260
aaggccatgt acgccccccc catecgccgc cagatccgct gcagcagcaa catcaccggc 1320
ctgctgctga cccgcgacgg cggcgaagag atcagcaaca ccaccgagt cttccgcccc 1380
ggcgcgccgg acatcgcgca caactggcgc agcgagctgt acaagtacaa ggtggtgaag 1440
atcgagcccc cggcgctggc cccaccatc gccatcaga cggtggtgga gagcgagaag 1500
agcgctgaga cctggggcgc catgttctgt ggtctctgtg gcgcggccgg cagcaccatc 1560
ggcgcccgca gctgacccct gaccgtgcag gcccgccagc tgctgagcgg catcgtlcaag 1620
cagcagaaca acctgctgcg cgccatcgag gcccgacagc acctgtgca gctgaccgtg 1680

```

```

ttggggcatca agcagctgca ggcgcgcgtg ctggccgtgg agcgtacct gaaggaccag 1740
cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccacgc cgtgccctgg 1800
aacgccagct ggagcaacaa gacgtcggac cagatcttga acaacatgac ctggatggag 1860
tgggagcggc agatcgacaa ctacaccaac ctgacttaca cctgatcga ggagagccag 1920
aacaccgagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
aactggttcg acatcagcaa gtggtgtgtg tacatctaac tcgag 2025

```

<210> 46
 <211> 1944
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp140.mut8.modSF162.delV2

```

<400> 46
gaattcccca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cagctcttctg ttctgcgccag cgcgtggagg aagctgtggg tgaccgtgta ctacggcgtg 120
cccggtgtgga aggaggccac caccacccgt ttctgcgccag gcgacgccaa ggctaccagg 180
accgaggttg ccaacgtgtg ggcacccacc gctgtgctgc ccaccgaccc caacccccag 240
gagatctgtc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtggg 300
cagatgcacg aggcacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccctgt gctgtaacct gactgcaccc aacctgaaga acgcccacaa caccagagc 420
agcaactgga aggaagatgga ccggcgccag atcaagaact gcagcttcaa ggtggggcgc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccaggg cctgcccacaa ggtgagcttc 540
gagcccatcc ccatccacta ctggcccccgc gccgggttgc ccatcctgaa gtgcaacgac 600
aagaagttca accgagcggg cccctgcacc aactgagca ccgtgcagtg caccacggcg 660
atccgcccgc tgggtgagcac ccagctgctg ctgaacggca gccgtggcga ggagggcgtg 720
gtgatccgca gcgagaactt caccgacaac gccaaagaca tcatcgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccgcccacac aacaacaccc gcaagagcat caccatcgcc 840
cccggccgcg cttcttaacg caccggcgac atcatcgccg acatccgcca ggcccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tctgacacca gctgcagggc 960
cagttcggca acaagaccat cgtgttcaag cagagcagcg ccggcgaccc ctagatcgtg 1020
atgcacagct tcaactcgcc cggcgagttc ttactatgca acagcaccac gctgttcaac 1080
agcacctgga acaacacat cggcccacac aacaccaacg gcacacatca cctgcctgac 1140
cgcatcaacg agatcatcaa ccgctggcag gaggtgggca aggcacatga cggccccccc 1200
atccgcccgc agatccgctg cagcagcaac atcaccggcc tgctgctgac ccgcgacggc 1260
ggcaaggaga tcagcaacac caccgagatc ttccgcccgc gggcgggcga catgcccagc 1320
aactggcgca gcgagctgta caagtacaag gtggtgaaga tcgagccctt gggcggtggc 1380
cccacacatc ccatcagcac cgtggtgacg agcgagaaga gcgcgtgac cctggcgccc 1440
atgttctcgg gcttctctgg cgcgcgcggc agcaccatgg gcgcccgcag cctgaccctg 1500
accgtgcagg cccgcagact gctgagcggc atcgtgcagc agcagaacaa cctgctgcag 1560
gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1620
ggcccgctgc tggccgtgga gcgctacctg aaggaccacg agctgctggc catctggggc 1680
tgcagcgcca agctgatctg caccacccgc gtgccctgga acgccaagct gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcca gatcgacca 1800
tacaccaacc tgatctacac cctgatcgag gagagcaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtgggcc agcctgtgga actggttcga catcagcaag 1920
tggctgtggt acatctaaat cgag 1944

```

<210> 47
 <211> 1836
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp140.mut8.modSF162.delV1/V2

<400> 47

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagttcttc ttctgcccag cgcgtggag aagctgtggg tgaccgtgta ctacggcggt 120
cccggtgtgga aggaggccac caccacctg ttctgcgcca gcgacgccaa ggctcatgac 180
accgaggtgc acaacgtgtg ggcacccac gctgtgctgc ccaccgaccc caacccccag 240
gagatcgctgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcacg aggcacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gctgtggcgc cggcaactgc cagaccagcg tgatcaccca ggctgtcccc 420
aaggtgagct tgaagcccat ccccatccac tactgogccc cgcccgctt cgcctacctg 480
aagtgcaacg acaagaagtt caacgcgagc ggcctgtgca ccaactgtgag cactgtgcag 540
tgaccacacg gcatcgcccc cgtggtgagc accagctgcg tgctgaacg cgactgtggc 600
gaggagggcg tggatgctcg cagcgagaac ttaccgcaca acgccaagac catcatcgtg 660
cagctgaagg agagcgtgga gatcaactgc accgcgcca acaacaacac cgcgaagagc 720
atccactatc gccccggcgc cgcttctac gccacggcg acatcatcgg cgactccgc 780
caggcccatc gcaacatcag cggcgagaag tggacaacaa cctgaagca gatcgtgacc 840
aagctgcagg ccagttctcg caacaagcac atcgtgttca agcagagcag cggcgcgac 900
cccgagctgc tgatgcacag ctcaactgc ggcggcgagt ttctctact caacagccac 960
cagctgttca acagcactgc gaacaacacc atcgccccca acaacaccaa cggcacccatc 1020
accctgccct gcgcctcaa gcagatcate acccgctggc aggaggtggg caagcgccatg 1080
tacgcccccc ccctcgcggc ccagatccgc tgcagcagca acatcacccg cctgtgtgct 1140
accgctgacg cggcgaagga gatcagcaac accaccgaga ttctccgccc cggcgggcgc 1200
gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtggtgaa gatcgagccc 1260
ctgggctgtg cccccacctc gcctacagc agcgtggtg agagcgagaa gagcgccgtg 1320
accctggggc ccattgttct gggcttcttg ggcgcgccc gcagcaccat ggcgccccgc 1380
agcctgaccc tgaccgtgca ggcccgcgag ctgctgagcg gcatcgtgca gcagcagaac 1440
aacctgctgc ggcgcacatg ggcgcagcag cactcgtgc agctgaccgt gtggggctac 1500
aagcactgctc aggcgcgctg gctggcgtgc gagcgtacc tgaaggacca gcagctgctg 1560
ggcagctggg gctgcagcgc caagctgac tgacccacgc ccgtgccctg gacgccagcc 1620
tgagacaaca agacctctga ccagatctgc aacaacatga cctgtgtgga gtggtgagcg 1680
gagatgcaca actacaccaa cctgatctac accctgatcg aggagagcca gaaccagcag 1740
gagaagaacg agcaggagct gctggagctg gacaaaggg ccagcctgtg gaactgtgtc 1800
gacatcagca agtggtctgt gtacatctaa ctcgag 1836

```

<210> 48

<211> 2547

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp160.modSF162

<400> 48

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagttcttc ttctgcccag cgcgtggag aagctgtggg tgaccgtgta ctacggcggt 120
cccggtgtgga aggaggccac caccacctg ttctgcgcca gcgacgccaa ggctcatgac 180
accgaggtgc acaacgtgtg ggcacccac gctgtgctgc ccaccgaccc caacccccag 240
gagatcgctgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcacg aggcacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gctgtggcgc cgcactgcac aactgaaga acgccaacaa cccaagagcc 420
agcaactgga aggagatgga cgcggcgagc atcaagaact cgagcttcaa ggtgaccacc 480
agcatcgcga acaaatgca gaaggagatc gccctgttct caaagctgga cgtgtgtccc 540
atcgacacga acaaccacag ctacaagctg atcaactgca acaccagctg gatcacccag 600
gctgccecca aggtgagctt cgagcccatc ccatccact actgcgccc cgccggtct 660
gccatcctga agtgcaacga caagaagttc aacggcgagc gccctgcac caactgtgagc 720
accgtgcagt gcacccacg gcctcgccc gtggtgagca cccagctgct gctgaacggc 780
agcctggcgc aggagggcgt ggtgatccgc agcgagaaat tcaccgacaa cgccaagacc 840
atcatcgtgc agctgaagga gagcgtggag atcaactgca cccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgccgc gcccttctac ccacggcgga catcatcgtg 960
gacatccgca aggcacactg caaacatcag ggcgagaagt ggaacaacac cctgaagcag 1020

```

atcgtgacca	agctgcagge	ccagttcggc	aacaagacca	tctgtttcaa	gcagagcagc	1080
ggcgcgacc	ccgagatcgt	gatgcacagc	ttaactcggc	gcggcgagtt	cttctactgc	1140
aacagcacc	agctgttcaa	cagcacctgg	aacaacacca	tccgcccaca	caacaccaac	1200
ggcaccatca	cctgcacctg	ccgcatcaag	cgatcatcat	accgtggcca	ggaggtgggc	1260
aaggccatgt	agcccccccc	catccggggc	cgatcccgct	gcagcagcaa	catcacccgc	1320
ctgctgctga	cccgcgacgg	cggcaaggag	atcagcaaca	ccaccgagat	cttccgcccc	1380
ggcgggcgcc	acatgcgcga	caactggcgc	agcgagctgt	acaagtacaa	gggtgtgaag	1440
atcagacccc	tgggctgtgc	ccccaccaag	gccaaagcgc	gcgtggtgca	gcgcgagaag	1500
cgcgctcgta	ccctggggcg	catgttctcg	ggcttctcgt	gcgcgcggcg	cagccacatg	1560
ggcgcccgca	gcttcgacct	gacgttgca	gcccggcagc	tgctgagcgg	catcgtcgag	1620
cagcagacaa	acctgctcgc	cgccatcgag	gccacagcag	acctgctgca	gatgcgcgtg	1680
tggggcatca	agcagctgca	ggcccgcgtg	ctggccgtgg	agcgctacct	gaaggaccag	1740
cagctgcgtg	gcactgtggg	ctgcagcggc	aagctgatct	gcaccaaccg	cgtgcccctg	1800
aacccgcagc	ggagcaacaa	gagcctggac	cgatctgga	acaacatgac	ctgtagggag	1860
tgggagcgcg	agatcgacaa	ctacaccaac	ctgatctaca	ccctgacgca	ggagagccag	1920
aaccagcagg	agaagaacga	gcaggagctg	ctggagctgg	acaagtggcg	cagcctgtgg	1980
aactgggttcg	acatcagcaa	gtggctgtgg	tacatcaaga	tcttcatcat	gatcgtgggc	2040
ggctgtgtgg	gcttcgcatc	cgtgttcacc	gtgctgagca	tctgtaaccg	cgtgcccagc	2100
ggctacagcc	ccctgagctt	ccagaccggc	ttccccgcc	cccgccggcc	cgaccgcccc	2160
gagggcatcg	aggaggaggg	cgccgagcgc	gaccgcgacc	gcagcagccc	cctggtgacc	2220
ggcctgctgg	ccctgatctg	ggacgacctg	cgacgctgtg	gcctgttcag	ctaccacggc	2280
ctgcgcgacc	tgatcctgat	cgccggccgc	atcgtggagc	tgctggggcg	cccgggctgg	2340
gagggcctga	agtactgggg	caacctgctg	cagtagctga	tccaggagct	gaagaaacag	2400
gccttgagcc	tgctcgacgc	catcgccatc	gcgctggccg	agggcacoga	cgtgatcatg	2460
ggaggtggcc	ccctcatcgg	ccgcgccttc	ctgcacatcc	cccgccgcat	cccgacaggc	2520
tgcagcgcg	ccctgctgta	actcgag				2547

<210> 49

<211> 2466

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp160.modSF162.delV2

<400> 49

gaattcgcca	ccatggatgc	aatgaagaga	gggctctgct	gtgtgctgct	gctgtgtgga	60
gcagtcttct	tttcgccag	cgccgtggag	aagctgtggg	tgacgtgta	ctacggcggt	120
cccggttgga	aggagccacc	caccaccctg	ttctgcggca	gcgacgcaca	ggcctacgac	180
accgaggtgc	acaacgctgt	ggccaccacc	gctgcgctgc	ccaccgaccc	caacccccag	240
gagatcgtgc	tggagaacgt	gaccgagaac	ttcaacatgt	ggaagaacaa	catgtgtggag	300
cagatgcacg	ggagcatcat	cagcctgtgg	gaccagagcc	tgaagccctg	cgtgaaagctg	360
acccccctgt	gcgtgacctc	gcactgcacc	aacctgaaga	acggcaccaa	caccaagagc	420
agcaactgga	aggagatgga	cgcgggcgag	atcaagaact	gcagcttcaa	ggtggggcgc	480
ggcgaactga	tcaactgcaa	caccagcgtg	atcaccacag	cctgcccaca	ggtgagcttc	540
gagcccatcc	ccatccacta	ctcgccccc	gcggcttcg	ccatcctgaa	gtgcaacggc	600
aagaagtcca	acggcagcgg	ccctgcacc	aacgtgagca	ccgtgcagtg	caccacaagg	660
atccgcccgc	tgttgagcac	ccagctgctg	ctgaacggca	gctggccgga	ggaggggctg	720
gtgatccgca	gcgagaactt	caccgacaa	gccaaagcca	tcatcgtgca	ctgcaaggag	780
agcgtgggga	tcaactgcac	ccgcccaccc	aacaacaccc	gcaagagcat	caccatcggc	840
cccgcccgcg	ccttctacag	caccggcgac	atcatcggcg	acatccgcga	ggcccacatgc	900
aacatcagcg	gcgagaagtg	gaacaacacc	ctgaagcaga	tctgtgacca	gctgcaggcc	960
cagtttcgca	acaagaccat	cgtgttcaag	cagagcagcg	gcggcgaccc	cgagatcgtg	1020
atgcacagct	tcaactgcgg	cgcgaggttc	ttctactgca	acagcaccga	cgtgttcaac	1080
agcactctga	acaacacatc	cggccccaac	aacacccaag	gcacatcaac	cctgcccctg	1140
cgcatacaag	agatcatcaa	ccgctggcag	gaggtgggca	aggcctatga	gcgcccccgc	1200
atccgcgccc	agatccgctg	cagcagcaac	atcacggccc	tgctgctgac	ccgcgacggc	1260
ggcaaggaga	tcaagcaaac	caccgagatc	ttccgcccgc	gcggcgcgga	catgcgcgac	1320

```

aactggcgca gcgagctgta caagtacaag gtgggtaaga tcgagccctt gggcggtggcc 1380
ccaccacaagg ccaagcgccg cgtggtgcag cgcgcgtgac cctggggcgcc 1440
atgttcctggt gttctcctggg cgcgcgcggc agcaccatgg cgcgcgcgag cctgaccctg 1500
accgtgcgag ccccgacagt gctgagcggc atcgtgcagg agcagaacaa cctgtgcgcg 1560
cccatcgagg cccagcagca cctgctgcag ctgacgctgt gggggatcaa cagctgcagc 1620
gcccgctgac tgcccgctga gcgctacctt aaggaccagg agctgctggg catctggggc 1680
tgacagcgga agctgatctg caccacgcgc gtgccttga acgccagctg gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgaga gatcgacaac 1800
tacaccaacc tgactctaac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtggggc agcctgtgga actggttoga catcagcaag 1920
tggctgtggt acatcaagat ctctacatgt atcgtggggc gctcgtggg cctgcgcatc 1980
gtgttctacc tgctgagcat cgtgaacggc gtgcgcggg gctacagccc cctgagcttc 2040
cagaccctgt tccccgcccc cgcgcgcccc gaccgcccc agggcatoga gggaggggc 2100
ggcgagcgcg acccgagacc cagcagcccc ctggtgcagc gctctgtgc cctgatctgg 2160
gacgacctgc gcagcctgtg cctgttcagc taccaccgce tgccgacct gatcctgat 2220
cgcgcccgca ctgtggagct gctggggcgc cgcgctggg aggcctgaa gtactggggc 2280
aacctgtgce agtactggat ccaggagctg aagaacagcg cctgagcct gtcgcagccc 2340
atcgccatcg cctgtggcga gggcaccgac cgcacatcg aggtggccca gcgcacggc 2400
cgcgctctcc tgcacatccc cgcgcgcac cgcagggct tcgagcgcc cctcgtgtaa 2460
ctcgag 2466

```

<210> 50

<211> 2358

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modSF162.delV1/V2

<400> 50

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgctgt gctgtgtgga 60
gcagttcttc ttgcgccag cgcctgtggg aagctgtggg tgaccgtgta ctacggcgtg 120
cccgctgtgga aggagggcac caccaccctg ttctgcgcca gcgacgcmaa ggccatcgac 180
accgaggtgc acacgtgtgt gggcaccaca gctcgtgctg ccaccgaccc caaccgccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggaacatcat cagcctgtgg gccagagccc tgaagccctg cgtgaacgtg 360
accctccctg gcgtggggcg cggcaactgc cagaccagcg tgatcaccca ggcctgcccc 420
aaggtgagct tcgagcccat ccccatccac taactgcgcc cgcgcggtt cgacatcctg 480
aagtgcacag acaagaagtt caacggcagc ggcccttgca ccaacgttag caccgtgcag 540
tgaccaccag cactccgccc cgtggtgagc acccagctgc tgctgaacgg caccgtggcc 600
gaggaggggc tgggtgatcg cagcgagaac ttaccgcaga accgcaagac catcatcgtg 660
cagctgaaag agagcgtgga gatcaactgc acccgcccc acaacaacac ccgcaagagc 720
atcaccatcg gccccggcgc cgccttctac gccaccggcg acatcatcgg cgacatccgc 780
caggcccaact gccaacatcag cggcgagaag tggaaacaaca ccttgaagca gatcgtgacc 840
aagctgcagg cccagttcgg caacaagacc atcgtgttca agcagagcag cggcgcgacg 900
cccgagatcg tgatgcacag ctccaactgc gggcgcgagt tcttctact caacagcacc 960
cagctgttca acagcacctg gaacaacacc atcgccccca acaacaacca cggcaccatc 1020
accctgcctt gcgcgatcaa gcagatcacc aacgcctggc aggggtggg caaggccatg 1080
tacgcccccc ccatccggcg ccagatccgc tgcagcaga acatcacggg cctcgtcgtg 1140
acccgcgacg cggcgaaagg gatcagcaac accaccgaga cttctcgccc cggcgcgccg 1200
gacatgcgcg acaactggcg cagcagagctg tacaagtaca aggtggtgaa gatcgagccc 1260
ctggcgctggt cccccacca ggccaagcgc cgcgtggtg agcgcgagaa gcgcgcgctg 1320
accctggggc ccatgttctt gggcttctct ggccgcggcg gcagcacat gggcgccgcg 1380
agcctgaccc tgacgctgca ggcccgccag ctgctgagcg gcactgtgca gcagcagaa 1440
aacctgctgc gcgcacatga ggccacgag caactgtgcy agctgacctg gtggggcgcc 1500
aagcagctgc agcccgctg gctggccgtg gagcgctacc tgaaggacca gcagctgctg 1560
ggcatcttgg gctgcagcgg caagctgac tcgaccacgc cctgcccctg gaacgcagcc 1620
tggagcaaca agagctgga ccagatctcg aacaacatga cctggatgga gtcgggagcg 1680

```

gagatcgaca	actacaccaa	cctgatctac	accctgatcg	aggagagcca	gaaccagcag	1740
gagaagaacg	agcaggagct	gctggagctg	gacaagtg	ccagcctgtg	gaactgggtc	1800
gacatcagca	agtggctgtg	gtacatcaag	atcttcac	tgatcgtggg	cgccctgggtg	1860
ggcctcgcca	tcgtgttcac	cgtgctgagc	atcgtgaacc	cggtgcgcga	gggctacagc	1920
ccctcgagct	tcagagcccg	cttccccgcc	ccccgcgcgc	ccgacccccc	cgaggcgcatc	1980
gaggaggagg	ggggcgagcg	cgacgcgcgc	cgacgcgcgc	ccctgggtgca	cggtcgtgctg	2040
gcccctgatct	gggacgacct	gcgcagcctg	tgctctgttc	gtaccacccg	cctgcgcgac	2100
ctgatcctga	tcgcgcgcgc	catcgtggag	ctcgtggggc	gccgcgctg	ggagggccctg	2160
aagtactggg	gcaactcgtc	gcagtaactg	atccaggagc	tgaagaacag	cgccgtggagc	2220
ctgtctcgag	ccatcgccat	cgcctggggc	gaggggcacc	accgcatcat	cgaggcggcc	2280
cagcgcctcg	gcgcgcctt	cctgcacatc	ccccgcgcga	tcgcgcaggg	cttcgagcgc	2340
gcccctgctgt	aactcgag					2358

<210> 51

<211> 1494

<212> DNA

<213> Human immunodeficiency virus

<400> 51

acaacagctct	tgtgggtcac	agtctattat	gggggtacctg	tgtggaaaga	agcaaccacc	60
actctgttttt	gtgcatcaga	tgctaaga	tacaaagcag	aggcacataa	cgctcgggct	120
acacatgcct	gtgtaccac	agaccccaac	ccacaggaag	taaatTTaac	aaatgtgaca	180
gaaattttta	acatgtggaa	aaataacatg	gtggaaacaga	tgcatgagga	tataatcagt	240
ttatgggcatc	aaagcctaaa	gccatgtgta	aaattaaccc	cactctgtgt	tacttttaat	300
tgtaactgata	agttgacagg	tagtactaat	ggcacaataa	gtactagtgg	cactaatagt	360
actagtggcca	cttaatgtac	tagtactaat	agtactgata	gttgggaaaa	ggtgcgcagaa	420
ggagaaataa	aaaactgctc	tttcaatc	accacaagt	taagagataa	agtcgagaaa	480
gaattattctc	tcttctataa	acttgatgta	gtaccaatag	ataatgataa	gtctagctat	540
agattgataa	attgtataac	ctcagtcatt	acacaagctc	gtccaaagg	atcttttgaa	600
ccaattccca	tacattattg	tgccccggct	ggtttttgca	ttctaagg	taaagataag	660
aagttcaatg	gaacaggacc	atgtaaaaat	gtcagcacag	tacaatgcac	acatggaatt	720
agaccagtag	tatcaactca	actgctgtta	aatggcagtc	tagcagaaga	agagatagta	780
cttagatctg	aaaattttac	agacatgtc	aaaaccataa	tagtcacagt	gaatgaattc	840
gtagaataa	attgtataag	acccaacaat	aatacaagaa	aaagtataca	tatatggacca	900
gggagagcat	tttatgcaac	agggtgataa	ataggagaca	taagacaagc	acattgttaac	960
attagttaa	caaaactggc	taacacttta	gaacagatag	ttgaaaaatt	aaagaaacaa	1020
tttgggaata	ataaaacaat	aatctttaat	tcatctctag	gagggggacc	agaaattgta	1080
tttccacagt	ttaattgtgg	aggggaattt	ttctattgta	atacatcaca	actatttaat	1140
agtaactgga	ataattactga	agagtgaaat	aagactaaa	aaatgtgac	tatcatactc	1200
ccatgcagaa	taagacaaat	tataaacatg	tggaagaag	taggaaaaag	aatgtatgcc	1260
cctcccatca	gaggacaat	taaatgttca	tcaaatatta	cagggctgct	ataactagaa	1320
gaggtgggta	ctaacaataa	taggacgaac	gacaccgaga	ccttcagacc	tgggggagga	1380
aaatgaagg	acaattggag	aagtgaatta	tataaatata	aagttagtaag	aattgaaaca	1440
ttaggagtag	caccacacca	ggcaaaagaa	agagtgtgtc	aaagagagaa	aaga	1494

<210> 52

<211> 2007

<212> DNA

<213> Human immunodeficiency virus

<400> 52

acaacagctct	tgtgggtcac	agtctattat	gggggtacctg	tgtggaaaga	agcaaccacc	60
actctgttttt	gtgcatcaga	tgctaaga	tacaaagcag	aggcacataa	cgctcgggct	120
acacatgcct	gtgtaccac	agaccccaac	ccacaggaag	taaatTTaac	aaatgtgaca	180
gaaattttta	acatgtggaa	aaataacatg	gtggaaacaga	tgcatgagga	tataatcagt	240
ttatgggcatc	aaagcctaaa	gccatgtgta	aaattaaccc	cactctgtgt	tacttttaat	300
tgtaactgata	agttgacagg	tagtactaat	ggcacaataa	gtactagtgg	cactaatagt	360
actagtggcca	cttaatgtac	tagtactaat	agtactgata	gttgggaaaa	gatgccagaa	420
ggagaaataa	aaaactgctc	tttcaatc	accacaagt	taagagataa	agtcgagaaa	480

```

gaattattcttc tcttctataa acttgatgta gtaccaatag ataattgataa tgcrtagctat 540
agattgtataa attgtaatac ctccagtcatt acacaagcct gtccaaaggt actcttttgaa 600
ccaatttccca tacattatttg tgcccgggct ggttttgcga ttctaaagtg taaagataag 660
aagtttcaatg gaacaggacc atgtataaaat gtccagcacag tacaatgcac acatggaatt 720
agaccagtag tatcaactca actgctgttta aatggcagtc tagcagaaga agagcagta 780
cttagattctg aaaaatttcc agacaatgct aaacccataa tagtacagct gaatgaactct 840
gtagaataa attgtataag acccaacaat aatacaagaa aaagtataca tataggacca 900
gggagagcat tttatgcaac aggtgatata ataggagaca taagacaagc acattgtaac 960
attagtaaa gcaactggac taacacttta gaacagatag ttgaaaaatt aagagaacaa 1020
tttgggaata ataaaacaat aatctttaat tcatctctag gaggggaccc agaaatttga 1080
tttcaagctt ttaactgtgg aggggaattt ttctattgta atacatcaca actatttata 1140
agtacctgga atattactga agaggtaaat aagactaaag aaaatgcac tatcatactc 1200
ccatgcagaa taagacaat tataaacatg tggcaagaag taggaaaagc aatgtatgcc 1260
cttcccatca gaggacaaat taaattgtca tcaaatatta cagggtctgt attaactaga 1320
gatgggtgta ctaacaataa taggacgaac gacaccgaga ccttcagacc tgggggagga 1380
aacatgaag acaattggag aagtgaaata tataaatata aagtagtaag aattgaacca 1440
ttaggagtag caccacccca ggcaagaga agagtgtgtc aaagagagaa aagacagtg 1500
ggactaggag cttgttctat tgggttcttg ggagcagcag gaagcactat gggcgacgcg 1560
tcagtgcagc tgacggtaga ggccagacaa tattgtctg tatalagtga acagcagaac 1620
aatttgcgtg aggtctattga ggcccaacag catctgtgtc aactcacggt ctggggcatc 1680
aaacagctcc aggcaagaat cctggtctgt gaaagatacc taaaggatca acagctccta 1740
gggatttggg ttgtctctgg aaaactcatt tgcaccataa ctgtgccttg aactcttagt 1800
tggagttaata aatctctgac tgagatttgg gataaatatg cctggatgga gtggaaaaga 1860
gaaatttggca atatacagg cttaatatag aatttaatg aaatagcaca aaacagacca 1920
gaaaagaatg aacaagaatt attggaatta gacaagtggt caagttttgtg gaatttgttt 1980
gatataacaa actgctgtg gtatata 2007

```

<210> 53

<211> 2532

<212> DNA

<213> Human immunodeficiency virus

<400> 53

```

acacacgtct tgtgggtcac agtctattat ggggtlacctg tgtggaaaga agcaaccacc 60
actctgtttt gtgcatcaga tgctaaagca tacaaagcag aggcacataa cgtctgggct 120
acacatgctt gtgtaccacc agaccccaac ccacaggagag taaatttaac aaattgtgaca 180
gaaaatttta acatgtggaa aaataacatg gtggaaacaga tgcattgagga tataactcag 240
ttatgggata aaagcctaaa gccattgtga aaattacccc cactctgtgt tacttttaaa 300
tgtactgata actgtgacagg tagtactaat ggcacaaata gtactagtgg actcaatagt 360
actagtggca ctaatagtac tagtactaat agtactgata ttgtggaaaa gatgcagaaa 420
ggagaataa aaaactgctc ttccaatc accacaaagt taagagataa agtagcagaa 480
gaatatcttc tcttctataa acttgatgta gtaccaatag ataattgataa tgcrtagctat 540
agattgataa attgtataac ctccagtcatt acacaagcct gtccaaaggt actcttttga 600
ccaattccca tacattattg tgcccgggct ggttttgcga ttctaaagtg taaagataag 660
aagtttcaatg gaacaggacc atgtataaaat gtccagcacag tacaatgcac actgggaatt 720
agaccagtag tatcaactca actgctgttta aatggcagtc tagcagaaga agagatagta 780
cttagactctg aaaaatttcc agacaatgct aaacccataa tagtacagct gaatgaactc 840
gtagaataa attgtataag acccaacaat aatacaagaa aaagtataca tataggacca 900
gggagagcat ttatgtcaac aggtgatata ataggagaca taagacaagc acattgtaac 960
attagttaa caaactggac taaacttcta gaacagatag ttgaaaaatt aagagaacaa 1020
tttgggaata ataaaacaat aatctttaat tcatctctag gaggggaccc agaaatttga 1080
tttcaagatt ttaattgtgg aggggaattt ttctattgta atacatcaca actatttaat 1140
agtcactgga atattactga agaggttaat aagactaaag aaatgacac tatcatactc 1200
ccatgcagaa taagacaat tataaacatg tggcaagaag taggaaaagc aatgtatgcc 1260
cttcccatca gaggacaaat taaattgtca tcaaatatta cagggtctgt attaactaga 1320
gatgggtgta ctaacaataa taggacgaac gacaccgaga ccttcagacc tgggggagga 1380
aacatgaag acaattggag aagtgaaata tataaatata aagtagtaag aattgaacca 1440
ttaggagtag caccacccca ggcaagaga agagtgtgtc aaagagagaa aagacagtg 1500
ggactaggag cttgttctat tgggttcttg ggagcagcag gaagcactat gggcgacgcg 1560

```



```

tcagtgcacg  tgacggtaca  ggccagacaa  ttattgtctg  gtatagtcca  acagcagaac  1620
aatttgcctga  gagctattga  ggccgaacag  catctgttgc  aactcacggt  ctggggcact  1680
aaacagctccc  aggcagaatt  cctggctgtg  gaaagatacc  taaaggatca  acagctccta  1740
gggatttggg  ttgtctctgg  aaaactcatt  tgacccaacta  ctgtgccttg  gaactctagt  1800
tggagtaata  aatctctgac  tgagatttgg  gataaatatga  cctggatgga  gtgggaaaga  1860
gaaatctggca  attatacagg  cttcaataac  aatttaattg  aaatagcaca  aaaccagcaa  1920
gaaaagaatg  aacaagaatt  attggaatta  gacaagtggg  caagtttgtg  gaattggttt  1980
gatataacaa  actggctgtg  gtatataaga  atattcataa  tgatagttag  aggcttgata  2040
ggtttaagaa  tagtttttgc  tgtactttct  atagtgaata  gagttaggca  gggatactca  2100
ccaatatcat  tgcagacccg  cctcccagct  cagagggggac  ccgacaggcc  cgaagggaatc  2160
gaagaagaag  tggagagag  agacagagac  agatccaatc  gattagtcca  tggattattg  2220
gcactcatct  gggacgatct  cgggagcctg  tgcctcttca  gctaccaccg  cttgagagac  2280
ttactcttga  ttgtagcgag  gattgtggaa  cttctgggac  gcagggggtg  ggaagccctc  2340
aagtatttgt  ggaatctcct  gcagtattgg  atcaggagac  taaagagtag  tgctgttagt  2400
ttgttaatg  ccacagcaat  agcagtagct  gaagggacag  ataggattat  agaataagta  2460
caaagaattt  ttagagctgt  aattcacata  cctagaagaa  taagacaggg  cttggagagg  2520
gctttactat  aa  2532

```

<210> 54

<211> 1599

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp120.modUS4

<400> 54

```

gaattcgcca  ccatggatgc  aatgaagaga  gggctctgct  gtgtgctgct  gctgtgtgga  60
cgagttcttg  ttctgcccag  cgcaccacc  gtgctgtggg  tgaccgtgta  ctacggcggt  120
ccggtgtgga  aggagggcca  caccaccctg  ttctgcgcca  gcgacgcca  ggcttacaag  180
cgcgagggcc  acaacgtgtg  ggccaccac  gctctgctgc  ccaccgacc  caaccgccag  240
gaggtgaacc  tgaccaacgt  gaccgagaac  ttcaacatgt  ggaagaacaa  catggttgag  300
cagatcgatg  aggacatcat  cagcctgtgg  gaccagagcc  tgaagccctg  cgtgaagctg  360
accccctgtg  cgtgtaccct  gaactgcacc  gacaagctga  ccggcagcac  caacggcacc  420
aacaagcaca  cgggcaccaa  cagcaccagg  ggccaccaaa  gcaccagcac  caacagcacc  480
gacagctggg  agaagatgcc  cgaggggcag  atcaagaact  gcagcttcaa  catcaccacc  540
agcgtgcgcg  acaaggtgca  gaaggagtac  agcctgttct  acaagctgga  cgtgggtccc  600
atcgacaacg  acaacgcca  ctaccgcctg  atcaactgca  acaccagct  gatcacccag  660
gcctgcccga  aggtgagctt  cgagcccatc  cccatccact  actgggcccc  cgcgggcttc  720
gccatctgca  agtgcaagga  caagaagtgc  aacggcaccg  gcccttgcaa  gaagctgagc  780
accgtgcagt  gcaccaccgg  catccgcccc  gtggtgagca  cccagctgct  cctggaacgg  840
agcctggcgc  agggaggatg  cgtgtcgcc  tcggagaact  tcaccgacaa  cgccaagacc  900
atcatctgct  agctgaacga  gtccgtggag  atcaactgca  tcgcgcccaa  caacaacagc  960
cgtgaagagca  tccacatcgg  ccccgcgcgc  gccttctacg  ccaccggcga  catcatcgcc  1020
gacatccgcc  agggccactg  caacatcagc  aaggccaaact  ggaccaacac  cctcgagcag  1080
atcgtggaga  agctgcggca  gcagttcggc  aacaacaaga  ccatcatctt  caacagcaga  1140
agcggcgccg  accccagatg  cgtgttccac  agcttcaact  cggcgggcga  gttctttcac  1200
tgcaacacca  gccagctgtt  caacagcacc  tggaaacatca  ccgagagggt  gaacaagacc  1260
aaggagacag  cacaccatcat  cctgccctgc  cgcattccgc  agatcatcaa  catgtggcag  1320
gaggtgggca  agggcattga  cgcgcccccc  atccggggcc  agatcaagt  cagcagcaat  1380
attaccggcc  ttctgctgac  ccgcagcggc  ggcaccaaca  acacgcgacc  caacgacacc  1440
gagaccttcc  gccccggcgg  cggcacaatg  aaggacaact  ggccgacgca  gctgtacaag  1500
tacaagggtg  tgcccatcga  gccctgggc  gtggccccca  cccaggccaa  gcgcccgcgt  1560
gtgcagcggc  aagaagccta  agatatcgga  tcctctaga  1599

```

<210> 55

<211> 1350

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp120.modUS4.del 128-194

<400> 55

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
gcagttcttg ttctgcccag cgcaccacc gtgtgtgtgt tgaccgtgta ctacggcggtg 120
cccggtgtgga aggaaggccac caccaccctg ttctgtgcca gcgacgcca ggcttacaag 180
gcggaggccc acaacgtgtg ggccaccac gccctgctgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaa ttcaacatgt ggaagaacaa catgtgtgag 300
cagatgcacg aggcacatcat cagcctgtgt gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt ggtgtggggg agggaaactgc gagaccagcg tgatcaccca ggctgtcccc 420
aagtgagct tcgagcccat ccccatccac tactgcgccc ccgcccgttt cgcctcctg 480
aagtgaacag acaagaagtt caacggcacc ggcccctgca agaactgtgag caccgtgcag 540
tgacccccag gcatccgccc cgtgtgtgag acccagctgc tctgtgaacg cagcctgccc 600
gaggaggaga tctgtgtgct ctccgagaa ttcccgaca accgcaagac catcatcgtg 660
cagctgaacg agtccgtgga gatcaactgc atccgcccc acaacaacac gcgtgaagag 720
ttcccatatc gccccggccc cgccttctac gccaccggcg acatcatcgg cgacatccgc 780
atggcccact gcaacatcag caaggccaac tggaccaaca cctctgagca gatcgtggag 840
aagctgtcgg agcagttcgg caacaacaag accatcatct tcaacagcag cagcggcgcc 900
gacccccaga tctgtgttca cagcttcaac tgcggcgccc agttcttcta ctgcaacacc 960
agccagctgt tcaacagcac ctggaacatc accgaggagg tgaacaagac caaggagagc 1020
gacaccatca tctctgccc cgcgatccgc cagatcatca acatgtggca ggaagtgagg 1080
aaggctcgtc atgccccccc catccggccc cagatcaagt gcagcagcaa tattaccggc 1140
ctgctgtcta ccccgacagg cggcaccaca aacaaccgca ccaacgacac cgagaccttc 1200
cgccccggcg gcggcaacat gaaggacaac tggcgacgag agctgtacaa gtacaagctg 1260
gtgcgcatcg agccccgtgg cgtggccccc acccaggcca agcgccgctg ggtgcagcgt 1320
gagaagcgct aagatatcgg atcctctaga

```

<210> 56

<211> 2112

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp140.modUS4

<400> 56

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
gcagttcttg ttctgcccag cgcaccacc gtgtgtgtgt tgaccgtgta ctacggcggtg 120
cccggtgtgga aggaaggccac caccaccctg ttctgtgcca gcgacgcca ggcttacaag 180
gcggaggccc acaacgtgtg ggccaccac gccctgctgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaa ttcaacatgt ggaagaacaa catgtgtgag 300
cagatgcacg aggcacatcat cagcctgtgt gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt ggtgtgccc gaactgcacc gaccagctga ccggcagcac caacggcacc 420
aacagcaca gcggcaccac cagcaccagc ggcaaccaac gcaccagcac caacagcacc 480
gacagctggg agaagatgcc cgaggcgag atcaagaact cgaacttcaa catcaacca 540
agcgtgtcgg acaagtgcca gaaggagatc agcctgttct acaagctgga cgtgtgtccc 600
atcgacaacg acaacgcccag ctaccgccc atcaactgca acaccagctg gatcacccag 660
gcctgcacca aggtgagctt cgagcccatc cccatccact actgcgccc ccgcccgttc 720
gcacactgga agtgcaagga caagaagttc aacggcacc ggcccctgca gaaactgagc 780
accgtgcagt gccaccacgg catccgcccc gtgtgtgaga cccagctgct gctgaacggc 840
agctcgtcgg aggaaggagat cgtgtgcgc tccgagaact tcccgacaa cgcgaagacc 900
atcatctgct agctgaacga gtccgtggag atcaactgca tccgcccaca caacaacacg 960
cgtaaagaca tccacatcgg ccccgcccgc gcttctaac ccaccggcga catcatcggc 1020
gacatccccc aggcaccctg caacatcagc aaggccaaact ggaccaacac cctcgagcag 1080
atcgtggaga agtgcggcga gcagtttcgg aacaacaaga ccatcatctt caacagcagc 1140
agcggcgccg accccgagat cgtgttccac agcttcaact cgccggcgga gtcttctcac 1200
tgcaacacca gccagctgtt caacagcacc tggaaatcga ccgaggaggt gaacaagacc 1260

```

```

aaggagaacg acaccatcat cctgccctgc cgcacccgcc agatcatcaa catgtggcag 1320
gaggtgggca agggcatgtg cgcctccccc atccgcggcc agatcaagtg cagcagcaat 1380
attaccggccc tgcctgctgac ccgcgaaggc ggcaaccaaa acaaccgcac caacgacacc 1440
gagacctctcc gccccggcgg cggcaacatg aaggacaact ggcgacgcga gctgtacaag 1500
tacaagggttg tgcgcacgca gccctctggg gtggccccc aacaggccaa gcgcgcgctg 1560
gtgcagcgcg agaagcgcgc cgtgggcccgt gggccctctg tcatcggtct cctggggccc 1620
gccgggagca ccatggggcg cgcctccgtg accctgacgc tgcaggcccg ccagctgctg 1680
agcgccatcg tgcagcagca gaacaacctg ctgcgcgcga tgcaggccca gcagcacctg 1740
tgcagctgca cgtgtgggg catcaagcag ctgcaggccc gcatcctggc cgtggagcgc 1800
taacctgaag accagcagct gctgggcatc tggggctgca gcggcaagct gatctgcacc 1860
accacctgac cctggaacag cagctgggag aacaagagcc tgaccagat ctgggaacac 1920
atgacctgga tggagtgga ggcggagatc ggcaactaca ccggcctgat ctacaacctg 1980
atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040
ggggccagcg ttgggaactg gttcgacatc aaccaatggc tgtggtacat ctaagatata 2100
ggatcctcta ga

```

2112

<210> 57

<211> 2112

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 gp140.mut.modUS4

<400> 57

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagctcttcg ttctgcgccag cgcaccaccg gtgctgtggg tgaccgtgta ctacggcggt 120
cccggtgtgga aggaggccac caccaccctg ttctgcgccg gcgacgcgcaa ggcttacaag 180
gccgagggccc caacagctgtg ggccaccacc gctctgctgc ccacgcgccc caacccccac 240
gaggtgaacc tgaccacaact gaccgagaac ttcaacatgt ggaagaacaa ctgtgtggag 300
cagatgcattg aggcacatcat cagcctgtgg gaccagagcc tgaagccctg ctggaagctg 360
acccccctgt gctgtgacct gaactgcacc gacaagctga ccggcagcac caacggcacc 420
aacagcacca gcggcaccaa cagcaccagc ggcaaccaaa gccaccgac caacagcaac 480
gacagctggg agaagatgcc cagggggcag atcaagaact gacgttcaa catcaccaac 540
agcgtgcgcg acaaggtgca gaaggagtag agcctgttct acaagctgga cgtggtgccc 600
atcgacaacg acaacgccag ctaccgcctg atcaactgca aacccagcgt gatcacccag 660
gcatctctga agtgcaagga caagaagttc aacggcaccc gccctgcaa gaacgtgagc 720
acgctgcagt caccaccagg catccgcccc gtggtgagca ccagctgctg gctgaacggc 780
agcctggccc aggaggagat cgtgctgcgc tcggagaact tcaccgacaa gcgcaagacc 840
atcatctctgc agctgaagga gtccgtggag atcaactgca tcgcccccaa caacaacacg 900
cgtgaagagca tccactcagg ccccgccgcg gccttctacg ccacccggga catcatcggc 1020
gacatccgcc aggccaccctg caacatcagc aaggccaact ggaccaacac cctcgagcag 1080
atcgtggaga agctgcgcga gcagttcgcc aacaacaaga ccatcatctt caacagacc 1140
agcggcgccg accccagatg cgtgttccac agcttcaact gcggcgcgga gttttctaac 1200
tgcaacacca gccagctgtt caacagcacc tggaaacatc ccgagggagt gaacaagacc 1260
aaggagaacg acccatcatc cctgccctgc cgcacccgcc agatcatcaa gctgtggcag 1320
gaggtgggca aggtccatgta cgcctccccc atccgcggcc agatcaagtc cagcagcaat 1380
attaccggcc tgcctgctac cgcgcagcgc ggcaaccaaa acaaccgcac caacgacacc 1440
gagacattcc gccccggcgg cggcaacatg aaggacaact ggccgacgca gctgtacaag 1500
tacaaggtgg tgcgcatcga gccctggggc gtggccccc aacaggccaa gcgcgcgctg 1560
gtgcagcgcg agaagagcgc cgtgggcccgt ggcgcctctg tcatcggtct cctggggccc 1620
gccgggagca ccatggggcg cgcctccgtg accctgacgc tgcaggcccg ccagctgctg 1680
agcggcatcg tgcagcagca gaacaacctg ctgcgcgcga tgcaggccca gcagcacctg 1740
ctcgagctga cgtgtgggg catcaagcag ctgcaggccc gcatcctggc cgtggagcgc 1800
taactgaagc accagcagct gctgggcatc tggggctgca gggaagcag gatctgcacc 1860
accacctgac cctggaacag cagctgggag aacaagaccc tgaccagat ctgggacaac 1920
atgacctgga tggagtgga ggcggagatc ggcaactaca ccggcctgat ctacaacctg 1980

```

```

atcgagatgc cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040
tgggccagcc tgtggaactg gttcgacatc accaactggc tgtggtacat ctaagatata 2100
ggatcctctca ga
2112

```

```

<210> 58
<211> 2181
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> Description of Artificial Sequence: gp140TM.modUS4

```

```

<400> 58
gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgccagc cgcaccaccg gtgctgtggg tgaccgtgta ctacgcgctg 120
ccgctgtgga aggaggccac caccaccctg ttctgtgccg gcgacgcgca ggcttacaag 180
gcggaggccc caaacgtgtg ggccaccacg gcctgcgtgc ccaccgaccc caaccgccag 240
gaggtgaacc tgaccaaact gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcatt aggcacatc cagcctgtgg gaccagagcc tgaagcctgt cgtgaagctg 360
acccccctgt gctgaccctt gaactgcacc gacaagctga ccggcgacac caacggcacc 420
aacagcacca cgggcaccaa cagcaccagc ggccaccaac gcaccagcac caacgacacc 480
gacagctggg agaagatgac cgagggcgag atcaagaact gcagcttcaa catcaccacc 540
agcgtgcgag acaaggtgca gaaggagtag agcctgttct acaagctgga cgtcggtgcc 600
atcgacaacg acaacgcccag ctaccgcctg atcaactgca acaccagcgt gatcaccag 660
gcctgcccga aggtgagctt cgagcccact cccatccact actgcccccc cgcggtcttc 720
gcatccttga agtgcaagga caagaagtgc aacggcaccg gccctgcaa gaactgtgagc 780
accgtgcagt gcacccacgg catccgcccc gtggtgagca cccagctgct gctgacaagg 840
agcctggccc aggaggagat cgtgctgcgc tcgagagaact tcaccgacaa cgccaagacc 900
atcatcgtgc agtgaacga gtccgtggag atcaactgca tcgcccccaa caacaacacc 960
cgtaaagaga tccacatcgg ccccgccgag gcttctatcg ccaccggcga catcatcgct 1020
gacatcgccc agggccactg caacatcagc aaggccaaat ggaccaaac cctcgagcag 1080
atcgttgaga agtgcgcgga gcagttcgcc aacaacaaga ccatcatctt caacagcagc 1140
agcggcgagg accccagatg cgtgttccac agcttcaact gcggcgcgga gttctcttac 1200
tgcaaaccca gccagctgtt caacagcacc tggaaatcca ccgagggagt gaacaagacc 1260
aaggagaacg acaccatcat cctgccctgc cgcattccgc agatcatcaa catgtgcgag 1320
gaggttggga agcccatgta cgcctccccc atccgcggcc agatcaagtg cagcagcaat 1380
attaccggcc tgctgtgacg ccgagcagcg ggcaccaaca acaaccgcac caacgacacc 1440
gagaccttcc gccccggcgg cggcaacatg aaggacaact gggcgacgga gctgtacaag 1500
tacaagttgg tgcgcactga gccctggggc gtggccccca cccaggccaa gcgcgcgctg 1560
gtgcagcgcg agaagcggcg cgtgggacct ggccctctgt tcacgtgctt cctgggcgag 1620
gccgggagca cctggggcgc cgcctccgtg accctgacgg tgcaggcccc cctgggcgag 1680
agcgcatcgc tgcagcagca gaacaacctg ctgcgcgcca tcgaggccca gcagcaccgt 1740
ctgcagctga ccgtgtgggg catcaagcag ctgcaggccc gcattcctggc cgtggagcgc 1800
tacctgaagg accagcagct gctgggcatc tggggctgca cgggcaagct gatctgcacc 1860
accacgtgac cctggaacag cagctggagc aacaagagac tgaccgagat cgtgggacca 1920
atgacctgga tggagtgagg gcgcgagatc ggcaactaca ccggcctgat ctacaacctg 1980
ctcagatcgc ccagaacca gcaggagaag aacgagcagg agctgctgga cgtggacaag 2040
tgggccaagc tgtggaactg gttcgacatc accaactggc tgtggtacat ccgcatcttc 2100
atcatgacg tggcgggcct gatcggcctg cgcactcgtg tcgctgtgct gagcatcgtg 2160
taagatcatg gatcctctag a
2181

```

```

<210> 59
<211> 1818
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> Description of Artificial Sequence:
gp140.modUS4.delV1/V2

```

<400> 59

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cgagctcttg ttctgccagc cgccaccacc gtgctgtggg tgaccgtgta ctaccgctgt 120
cccgctggga aggaggccac caccaccctg ttctgcgcca gcgacgccaa ggcctacaag 180
gcggaggccc acaactgtgt ggcaccaccac gctctgctgc ccaccgaccc caacccccag 240
gaggtgaacc tgaccaactg gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcgat aggacatcat cagcctgtgg gaccagagcc cccatccccca tccactactg cgcccccgcc 420
ggccaaggct gcccacaggt gacgttccag cccatccccca tccactactg cgcccccgcc 480
gtgagcaccg tgcagtgac ccacggcctc cgcccccgct tgagacccca gctgctgctg 540
aacggcagcg tggccgagga ggagatcgtg ctgctgctgc agaacttacc gcacaacgcc 600
aagaccatca tctgtgcagt gaacgagctc gtggagatca actgcatcgc ccccaacaac 660
aacacgcgta agagcatcca catcgccccg gcccgcgcct tctacgccac cgccgacatc 720
atcgccgaca tcccgaccg ccaactgaac atcagcaagg ccaactggac caacaccctc 780
gagcagatcg tggagaagct gcgcgagcag ttccggaaca acaagaccat catcttcaac 840
agcagcagcg gcggcgacc cgagatcgtg ttccacagct tcaactgcgg cgccgagttc 900
ttctactgca acaccagcca gctgttcaac agcacctgga acatcacga ggaggtgaa 960
aagaccaggg agaaccagac catcatcctg cctcgcccca tcccgccagat catcaactat 1020
tggcagaggg tgggcaagcg catgtacgcc ccccccatcc gcggccagat caagtgcagc 1080
agcaatatta ccggcctgct gctcagccgc gacggcgcca ccaacaacaa ccgcaccaac 1140
gacaccgaga ccttcgcccc cggcgccggc aacatgaagg acaactggcg cgccgagctg 1200
tacaagtata agtggtgtgc catcgagccc ctgggctgtg cccccaccga ggcgaagcgc 1260
cgctgtgtgc agcgcgagaa gcgcgcctgt ggcctggggc cctgtttcat cggtctctgt 1320
ggcgccgccc ggagcaccat gggcgccccc tccgtgaccc tgaccgtgca ggcgcgcag 1380
ctgtgagcgc gcatcgtgca gcagcagacc aacctgctgc gcgccatcga ggccacagca 1440
cacctgtctc agctgacctg gtggggctatc agcagctgct ggcctctggg gctgcagcgg caagtgtcat 1500
tgacgctacc cagtgaccga gcagctgctg tggagcaaca agagcctgca cagagatctg 1560
tgacacacca cgtgccccg gaacagcagc tggagcaaca agagcctgca cagagatctg 1620
gacacatgta cctgtgtgga gtgggagcgc gagatcgcca actacaccgg cctgatctac 1680
aacctgacgc agatcgccca gaaccagcag gagaagaacy agcaggagct gctggagctg 1740
gacaagtggg ccagcctgtg gaactggttc gacatcacca actggtctgt gtacatctaa 1800
gatatcggat cctctaga 1818

```

<210> 60

<211> 2031

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 gp140.modUS4.delV2

<400> 60

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cgagctcttg ttctgccagc cgccaccacc gtgctgtggg tgaccgtgta ctaccgctgt 120
cccgctggga aggaggccac caccaccctg ttctgcgcca gcgacgccaa ggcctacaag 180
gcggaggccc acaactgtgt ggcaccaccac gctctgctgc ccaccgaccc caacccccag 240
gaggtgaacc tgaccaactg gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcgat aggacatcat cagcctgtgg gaccagagcc tgaggccctg cgtgaagctg 360
acccccctgt cgtgacctc gaactgcacc gacaagctga ccgcgagcac caacggcacc 420
aacagcacca gcggcaccga cagcaccagc ggccaacaa gcaccagcac caacagcacc 480
gcagcctggg agaagatgcc cgaggcgag atcaagaact gcagcttcaa cctggcgccc 540
ggccgctgta tcaactgcga cccagcgtg atcaaccagg cctgccccaa ggtgagcttc 600
gagcccatcc ccatccacta ctgcgccccg gccggtcttc ccatcctgta gtgacaagg 660
aagaagttca agggcaccgg cccctgcga aacgtgagca ccgtgcagtg caccacggcg 720
atccgccccg tggtagcac ccagctgctg ctgaacggca gcttggcgga ggagagagatc 780
gtgctgcgct ccgagaactt acccgacaac gccaaagaca tcatcgtgca gctgacaagg 840
tccgtggaga tcaactgcat ccgccccacc aacaacacgc gtaagagcat ccaactcgcc 900
cccgccgcyg cctttacgca caccggcagc atcagcgggc acatcgcca ggccactcgc 960

```

```

aacatcagca aggcacactg gaccaacacc ctgcagcaga tcgtggagaa gctgcgcgag 1020
cagttcggcca acaacaagac catcatcttc aacagcagca gcgcgcgcga ccccgagatc 1080
gtgtttccaca gtttcaactg cggcggcgag ttctttact gcaacaccag ccagctgttc 1140
aacagacactt ggaacatcac cgaggaggtg aacaagacca agggagaacga caccatcatc 1200
ctgcctctgcc gcatccgccca gatcatacaac atgtggcagg aggtggggcaa ggccatgtac 1260
ccccccccca tccgcggcca gatcaagtgc agcagcaata ttaccggcct gctcgtgacc 1320
cgcgacggcg gcaccaacaa caaccgcacc aacgacacc agaccttcg ccccggcggc 1380
ggcaacatga aggacaactg gcgcagcgag ctgtacaagt acaaggtggt gcgcacatga 1440
ccccggggcg tggccccac ccaggccaag cgcgcgtgct tgcagcgcga gaagcgcgcg 1500
gtggggcttgg gcgccttgtt catcggtctc ctggggcgcc cggggagcac catggggcgc 1560
gcctcctgta ccttgacctg gcaggcccg cagctgctga gcgcgcatgt gcagcagcac 1620
aacaacctgc tgcgcgccat cgaggccag cagcactctg tgcagctgac cgtgtggggc 1680
agctggagca acaagagcct gaccgagatc tgggacaaca tgacctggt ggagtgggag 1740
ctgggcatct ggggctgcag cggcaagtgc atctgcacca ccacctgctc ctggaacagc 1800
agctggagca acaagagcct gaccgagatc tgggacaaca tgacctggt ggagtgggag 1860
cgcgagatgc gcaactacac cggcctgata tacaacctga tcgagatcgc ccgaaaccag 1920
caggagaaga acgagcagga gctgctggag ctggacaagt gggccagcct gtggaactgg 1980
ttcgacatca ccaactggct gtgtgtacat taagatatgc gatcctctag a 2031

```

<210> 61

<211> 1818

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.mut.modUS4.delV1/V2

<400> 61

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgctgt gctgtgtgga 60
gcagttcttc tttcgccagc cggccaccac gtgctgtggg tgacctgtga ctacggcgtg 120
cccgctgtgga aggagggcac caccaccctg ttctgcgcga gcgacgcaa ggcttacaag 180
gccgagggccc acaacgtgtg ggccaccacc gctcgtgctg ccaccgacc ccacccccag 240
gaggtgaacc tgaccaactg gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcagt aggcacatcat cagcctgtgg gaccagagct tgaagccctg cgtggggcgc 360
ggccaggcctt gcccaagggt gagcttcgag cccatcccca tccactact gcgcccgcc 420
ggcttcgcca tctgaagtgc caaggacaag aagttcaag gcaccggccc ctgcaagaac 480
gtgagcaccg tggccyagga ggagatcgtg ctgcgctccg agaacttcac cgacaagccc 540
aagaccatca tgcgtgcagt gaacgagctc gtggagatca actgcactcg ccccaacca 600
aacacgcgta agagcatcca catcgccccc ggccgcgctc tctacgccac cggcgacatc 720
atcgcgacga tgcgccaggg ccactgcaac atcagcaagg ccaactggac caacaccctc 780
gagcagatcg tggagaagct gcgcgagcag ttccgcaact acaagacct catcttcaac 840
agcagcagcg gccggcagccc cgagatcgtg ttccacagct tcaactcggc cggcgagttc 900
ttctactgca accaccgcca gctgttcaac agcacctgga acatcacgga ggaggtgaac 960
aagaccaagg agaagcagac catcatcctg cctgcgcga tcccgagat catcaactgt 1020
tggcaggagg tgggcaaggg catgtacgcc ccccaccatc gcggccagat caagtgcagc 1080
agcaaatata cggcctgctg gctgacccgc gacggcgcca ccaacaacaa ccgcaacca 1140
gacaccgaga ccttcgcgcc cggcgccggc aacatgaagg acaactggcg cagcgagctg 1200
tacaagtaca agtgggtggc catcgagccc ctgggctggt cccccacca ggccaagcgc 1260
agcgtgtggc agcgcgagaa gagcgcgctg gcctggggcg cctgttcat ggcttctctg 1320
ggcgccgcgc ggagcaccat ggccgcgcgc tccgtgacc tgcacctgca ggcccgccag 1380
ctgctgagcg gcatcgtgca gcagcagaac aactctgtgc gcgccatoga ggccaccagc 1440
cacctgtgct agctgacctg gtggggcalt aagcagctgc aggcccgat agggccgctg 1500
gagcgtacc tgaaggacca gcagctgtgt ggcacttggg gctgcagcgg caagctgate 1560
tgacacacca ccgtgacctg gaacagcagc tggagcaaca agagcctgac cagatctgtg 1620
gacaacatga ccttgatgga gtggggagcg gagatcgcca actacaccgg cctgatctac 1680
aacctgatcg agatcgccca gaaccagcag gagaagaacg agcaggagct gctgttctgt 1740
gacaagtggg ccagcctgtg gaactgggtc gacatcacca actggctgtg gtacatctaa 1800

```

gatatcggtat cctctaga

1818

<210> 62

<211> 1818

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.modUS4.del 128-194

<400> 62

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cgagttcttc ttctgcccag cgcaccacc gtgctgtggg tgaccgtgta ctacggcggt 120
ccggtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcca ggctttacaag 180
cccgaggccc acaacgtgtg ggcaccacc gctgtgctgc ccaccgaccc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatcgcat aggcacatcat cagcctgtgg gaccagagcg tgaagccctg cgtggggccc 360
ggccaggccct gccccaaggt gagcttcgag cccatcccca tccactactg cgcccccgcc 420
ggcttcgcca tcttgaagtg caaggacaag aagttcaacg gcaccggccc ctgcaagaac 480
gtgagcaccc tgcagtgcac ccacggcctc cgccccgtgg tgagcaccca gctgtgctgt 540
aacggcgacc tggccgagga ggagatcggt ctgcgctccg agaacttcaac cgacacggcc 600
aagaccatca tctgtcagct gaacgagtcg gtggagatca actgcatccg ccccaacaac 660
aatccgcgta agagcatcca catcgcccc ggccgcgctc tctacgccac cggcgacatc 720
atcgccgaca tccgcaaggc ccaactgcaac atcagcaagg ccaactggac caacacccct 780
gagcagatcg tggagaagct gcgcgagcag ttccgcaaca caaagaccaa catcttcaac 840
agcagcagcg cggcgacacc cgagatcggt ttccacagct tcaactcgcg cggcgagctc 900
ttctactgca acaccagcca gctgttcaac agcaccttga acatcacga ggaggtgaac 960
aagacaagg agaacgacac catcatcctg cctgcgcga tccgcagat catcaacatg 1020
tgccaggagg tgggcaaggc catgtacgcc ccccccatc cggcgagat caatgacagc 1080
agcaatatta cggcctctgt gctgaccgcc gacggcgcca ccaacaacaa ccgcaccaac 1140
gacaccgaga ccttcggccc cggcgggcgc aacatgaagg acaactggcg cagcgagctg 1200
tacaagtaca aggtggtgct catcgagccc ctggggcggt cccccacca ggccaagcgc 1260
cgcggtgtgt agcgcgagaa gagcgccgtg ggcctggggc cctgttctat cggcctcctg 1320
ggcgccgcgc ggagcaccat ggcgcgcgcg tccgtgacc tgaccgtgca ggcccgcagc 1380
ctgctgagcg gcactgtgca gcagcagaac aaactgctgc gcgccatcga ggcccagcag 1440
cactgtgctc agctgacgct gtggggcctc aagcagctgc agggcccgcat cctggccgtg 1500
gagcgctacc tgaaggacca gcagctgctg ggcattctgg gctgcagcgg caagctgac 1560
tgaccaccca cgtgcccctg gaacagcagc tggagcaaca agagcctgac cgagatctgg 1620
gacaacatga cctggatgga gtgggagcgc gagatcgcca actacacggg cctgatcatc 1680
aaactgatcg agatcgcccc gaaccagcag gagaagaacg agcaggagct cgtggagctg 1740
gacaagtggg ccagcctgtg gaactgtgtc gacatcacca actggctgtg gtacatctaa 1800
gatatcggtat cctctaga
1818

```

<210> 63

<211> 1863

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.mut.modUS4.del 128-194

<400> 63

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cgagttcttc ttctgcccag cgcaccacc gtgctgtggg tgaccgtgta ctacggcggt 120
ccggtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcca ggctttacaag 180
gccgaggccc acaacgtgtg ggcaccacc gctgtgctgc ccaccgaccc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300

```

```

cagatgcgatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgggggc agggaaactgc gagaccagcg tgatcaccca ggccctgccc 420
aaggtgagct tccagcccat ccccatccac tactgcgccc ccgcggctt cgcactcctg 480
aagtgaagct tcaagaagtt caacggcacc ggcccttgca agaactgag caccgtgcag 540
tgaccaccag gcatccgccc cgtggtgagc accagctgct tgctgaacgg cagcctggcc 600
gaggaggaga tegtgtgtgc ctccgagaac ttaccgcga acgcgaagac catcatcgtg 660
cagctgaacg agtccgtgga gatcaactgc atccgcccc acaacaacac gcgtgaagac 720
atccacatcg gccccggcgc cgccttctac gccaccggcg acatcatcgg cgacatccgc 780
caggccccact gcaacatcag caaggccaac tggaccaca cctcagca gaictgtggag 840
aagctgcgag agcagttcgg caacaacaag accatcatct tcaacagcag cagcggcgcc 900
gaccgccaga tctgtgtcca cagcttcaac tgcggcgccg agttcttcta ctgcaacacc 960
agccagctgt tcaacagcag ctggaacatc accgaggagg tgaacaagac caaggagaac 1020
gacaccatca tctctgccctg ccgcatccgc cagatcatca acatgtggca ggaggtggcg 1080
aaggccatgt acgccccccc catccgcgcc cagatcaagt gcagcagcaa tattaccggc 1140
ctgctgtcga cccgcgacgg cggcaccaac aacaaccgca ccaacgacac cgagaccttc 1200
cgccccggcg cggcgcaact gaaggcaaac tggcgacggc agctgtacaa gtacaaaggt 1260
gtgcgcatcg agccctctgg cgtggccccc acccaggcca agcccgcgct ggtgcagcgc 1320
gagaagagcg ccgtgggctct gggcgccctg ttcactggct tctgtggcgc ccgcggagcg 1380
acctggggcg ccgctccgtg gacctgacc gtgcaggccc gccagctgct gagcggcgtc 1440
gtgcgacgag agaacaacct gctgcgcgcc atcgaggccc agcagcaact gctgcagctg 1500
accgtgtggg gcatcaagca gctgcagccc cgcactcttg ccgtggagcg ctaccagtag 1560
gaccagcagc tgtctgggcat ctggggctgc agcggcaagc tgatctgcac caccacgctg 1620
ccctggaaca gcagctggag caacaagagc ctgaccgaga tctgggacaa catgacctgg 1680
atggagtggg agcgcgagat cggcaactac accggcctga tctacaacct ctgcagagct 1740
gcccgagaac agcaggagaa gacagagcag gagctgctgg agctggacaa gtggggcagc 1800
ctgtggaaat ggttcgacat caccaactgg ctgtggtaca tctaagatat cggatcctct 1860
aga

```

<210> 64

<211> 2634

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp160.modUS4

<400> 64

```

gaattcgcca coatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
gcagtctctg tttcgcgccg cggcaccacc gtgtgtgtgg tgaccgtgta ctacggcggt 120
cccggtgtga aggagggcac caccaccctg ttctggtgca gcgacgcgaa ggcttacaag 180
ggcagagccc acaactgtgt ggccaccacc gctcgtgtgc ccacgcagcc caaccaccag 240
gaggtgaacc tgaccaactg gaccgagaac ttcaactatg ggaagaacaa catggtggag 300
cagatgcgat aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgacctt gaactgcacc gacaagtga ccggcagcac caacggcacc 420
aacagcacca gcggcaccaa cagcaccagc ggccaacaac caccagcac caacaccacc 480
gacagctggg agaagatgcc cgagggcgag atcaagaact cgagcttcaa catcacccacc 540
agcgtgcgcg acaaggtgca gaaggagtac agcctgttct acaagctgga cgtgtggccc 600
atcgacaagc acaacgccag ctaccgctgt atcaactgca acaccagctg gatcacccag 660
gctcgcacca aggtgagctt cgagcccatc cccatccact actgcgcccc cgccgcttct 720
gccactctga agtgcaagga caagaagttc aacggcacgg ccctctgcaa gaactgtgag 780
acgtctgagc gacaccacgg catccgcccc gtggtagaga cccagctgct gctgaacggc 840
agcctggcgg aggaggagat cgtgctgcgc tccgagaaat tcaccgacaa gcccaagacc 900
atcatctgtc agctgcaacga gtcctggagg atcaactgca tccgcgccaa caacaaccag 960
cgtaaagaca tccacatcgg ccccgccgcg gccctctacg ccaccggcga catcatcggc 1020
gacatccgce agggccactg caacatcage aaggccaact ggacccaac cctcagagcg 1080
atcgtggaga agtgcgcgca gcaagtccgg aacaacaaga agcttcaact gcggcgcgga gttcttctac 1200
agcggcgggc accccagagat cgtgtttcac agcttcaact tgaagaatca ccgaggaggt gaacaagacc 1260
tgcaacacca gccagctgtt caacagcaga cgtatccgce agatcatcaa catgtggcag 1320
aaggagaacg acacatcatc cctgcccctg

```



```

gaggtgggcca agggccatgta cgcgcgcgcgc atccgcgggccc agatcaagtg cagcagcaat 1380
attaccggccc tgcctgtgac cgcgcgcgcgc ggcaccaaca acaaccgcac caacgcacac 1440
gagacccctcc gcccccgcgg cgcgcacatg aaggacaaat ggcgcgcgca gctgtacaag 1500
tacaaggtgg tgccgatcga gccctcgggc gtggccccc cccaggccaa gcgcgcgctg 1560
gtgcagcgcgc agaagcgcgc cgtgggcctg ggcgcgcctg tcatcggtt cctgggcgcgc 1620
gccgggagcga ccatgggcgc gcctcgcgtg accctgacgc tgcaggccgc ccagctgctg 1680
agcgcgcctgc tgcagcagca gaacaacctg ctgcgcgcgc tcgaggccca gcagcaccctg 1740
ctgcagctgca ccgtgtgggg catcaagcag ctgcaggccc gcactcctgg cgtggagcgc 1800
tacctgaagg accagcagct gcctgggcac aacaagagcc tgaccgcgat ctgggacaac 1920
atgaccttggg tggagtggga gcgcgagatc ggcacactaca ccgcctgat ctacaacctg 1980
atcgagatgc cccagaaaca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040
tgggcacgccc tgtggaactg gttcgacatc accaactggc tgtggtacat ccgcatcttc 2100
atcatgatgc tgggcggcct gatcggcctg cgcactcgtg tcgcgctgct gagcatcgtg 2160
aacgcgcgtgc aacagggcta cagccccatc agcctgcaga ccgcctgcc ccccccgcgc 2220
ggcccgcgacc gccccgaggg catcgaggag gaggggcgcc agcgcgacgc cgaccgcgcg 2280
aacgcgcctgg tgcacgcctg gctggccctg atctgggagc acctgcgcga ctctgcctg 2340
ttcagctacc accgcctcgc gcacctcgtg ctgactcgtg ccgcactcgt ggagctcgtg 2400
ggcgcgcgcgc gctgggaggg cctgaagtac tgggtggaac tgcgcagta ctggagccag 2460
gagctgaaga gcacgcgcgt gagcctgttc aacgccacgc ccactcgctg ggcgcggggc 2520
accgcagcga tcactcgagat cgtgcagcgc atctccgcgc cgtgatcca catccccgc 2580
cgcctccgccc agggcctgga gcgcgcgcctg ctgtaagata tcggatcttc taga 2634

```

<210> 65

<211> 2538

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modUS4.delV1

<400> 65

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cgagtcttgc ttctgcccg cgcacaccac gtgctgtggg tgaccgtgta ctaccgctgt 120
ccgctgtgga aggaggccac caccaccctg ttctgcgcga gcgacgccaa ggccttacaag 180
gcgcaggccc acacagctgt ggcacccacc gccctgcgtgc ccaccgaccc caaccgccag 240
gaggtgaacc tgaccaacgt gaccgagaaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcgatg aggcacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gaactgcacc gacaagctgg gcgcgcggcg cgaagatcaa 420
aactgcagct tcacacatcac caccagcgtg gcgcacaagg tgacagaagg tcacagcctg 480
ttctacaagc tggagctggt gcccatcgac aacgacaaag ccagctaccg cctgatcaac 540
tgcaaacaca gcgtgatcac ccaggcctgc cccaaggtga gcttcgagcc catccccatc 600
cactactgcg ccccgcgcgg ctctgccatc ctgaagtga aggaacaaga gttcaacggc 660
accggccctc gcagaagcgt gagcacctgt cagtgccacc ccggtcgtgt 720
agcaccaccg tctgtctgaa cggcagcctg gccgaggagg agactcgtgt gcgctccgag 780
aacttcaacc acaacgccaa gaccatcatc gtgcagctga acgagtcctg ggagatcaac 840
tgcatccgcc ccaacaacaa cagcgtgaag agcatccaca tcggcccccgc cgcgcctctc 900
taagccaccg gcgacatcat cggcgacatc cgcacggccc actgcaacat cagcaaggcc 960
aactggacca acactcctga gcagatcgtg gagaagctgc gcgagcagtt cgcgacaacac 1020
aagaccatca tcttcaacag cagcagcggc ggcgaccccg agatcgtgtt ccaacagcttc 1080
aactgcggcg gcgagttctt ctactgcaac accagccaga tgttcaacag cactctggaac 1140
atcaccaggg aggtgaacaa gaccaaggag aacgacacca tcatcctgcc ctgcgcgcat 1200
gcgcagatca tcaacatgtg gcaggaggtg ggcaaggcca tgcacgcccc ccccatccgc 1260
ggccagatca agtgcagcag caatatcacc gcctcgtgct tgaccgcgca cggcggcacc 1320
aacaacaacc gcaccaacga caccgagacc ttccgccccg gcggcgccca catgaaggac 1380
aactggcgca gcgagctgta caagtacaag gtggtgcgca tcgagccccc gggcgtggcc 1440
cccaccaggc ccaagcgcgc cgtggtgcag gcgcgagaag gcgcgctggc cctggggcgc 1500
ctgttcatcg gcttcctggg gcgcgcgcgc agcaccatgg gcgcgcgcct cgtgaccctg 1560

```

```

accgtgcagg cccgcagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1620
gccatcgagg ccacgacgca cctgctgcag ctgaccctgt ggggcatcaa gcagctgcag 1680
gcccgcatcc tggccgtgga gcgtacctgt aaggaccagg agctgctggg catctggggc 1740
tgcagcgcca agctgatctg caccaccacc gtgcccctga acagcagctg gagcaacaag 1800
agcctgaccg agatctggga caacatgacc tggatggagt gggagcgcca gatcggaac 1860
tacaccggcc tgatctacaa cctgategag atcgcccaga accagcagga gaagaacagag 1920
caggagctgc tggagctgga caagtggggc agcctgtgga actggttcga catcaacca 1980
tggctgtggt tgcctgagca ctctcatcat atcgtggggc gctgatctgg cctgcgcctc 2040
gtgttcgcgc tgcctgagca ctgtaaacgc gtgcgcaggg gctacagccc catcagcctg 2100
cagaccgcgc tgcgcgcaca gcgcggccgc gacgcgcgcg agggcatcga ggaagagggc 2160
ggcagcgccg acccgacagg cagcaaacgc ctggtgcagc gctgctggc cctgatctgg 2220
gacgacctgc gcagcctgtg cctgttcagc taccaccgcc tgcgcgacct gctgctgac 2280
gtggcccgca tctgtggagt gctggggcgc cgcgctggg aggcctgaa gctgctgac 2340
aacctgctgc agtactggag ccaggagctg aagagcagcg ccgtgagcct gttcaacgac 2400
accgcctatc ccgtggcgga gggcaccgac cgcctcatcg agatcgctga gcgcattctc 2460
cgcgcgctga tccacatccc ccgcgcgcat cgcaggggcc tggagcgcgc cctgctgtaa 2520
gatatcggat cctctaga

```

<210> 66

<211> 2553

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 gp160.modUS4.delV2

<400> 66

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cgagttcttc ttctgcccag gcgcaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccggtgtgga aggaggccac caccaccctg ttctgcccga gcgacgcgaa ggcttacaag 180
gccgagggccc acaactgtgt ggccaccacc gcctgcgtgc ccaccgaccc caaccccacg 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcgat gagacatcat cagcctgtgg gaccagacc tgaagccctg ctgtaagctg 360
acccccctgt gcgtgacctt gaactgcacc gacaagctga ccggcagcac caacggcacc 420
aacagcacca cgggcaccaa cagcaccagc ggcaaccaaa gcaccagcac caacagcacc 480
gacagctggg agaagatgcc cgagggcgag atcaagaact gcagcttcaa catcgggccc 540
ggcgcctgca tcaactgcaa caccagctgt atcaccacag cctgcccaca ggtgagcttc 600
gagcccatcc ccatccacta ctgcgccccc gcgggcttcg ccatcctgaa gtgcaaggac 660
aagaagtcca acggcaccgg cccctgcaag aacgtgagca ccgtgcagtg caccacggc 720
atcgcgcccg tggtagacac ccagctgctg ctgaacggca gcctggcgga ggaggagatc 780
gtgctgcgct cagagaactt caccgcaaac gccaaagaca tcatcgtgca gctgaacgag 840
tcctgtgaga tcaactgcat ccgcgccaac aacaacacgc gtaagagcat ccacatgcc 900
cccggcgcg ccttctacgc caccggcgac atcatcgccg acatccgcca ggcccactgc 960
aacatcagca agccacaactg gaccaaaccc ctcgagcaga ctcgggagaa cctgcgcgag 1020
cagttctggca acacaagac catcatcttc aacagcaga cggcgggcga ccccgagatc 1080
gtgttccaca gcttcaactg cggcgggcag ttcttctact gcaacaccag ccagctgttc 1140
aacagcaact ggaacatcac cgaggagggt aacaagacca aggagaacga caccatcac 1200
ctgcgctgcc gcatcgcgca gatcatcaac atgtggcagg aggtgggcaa ggccatgtac 1260
gcccccccca tccgcgcca gatcaagtgc agcagcaata ttaccggcct cctgctgacc 1320
cgcgacggcg gcaccaacaa caaccgcacc aacgacacgc agaccttccg ccgcgcgcgc 1380
ggcaacatga aggacaactg gcgcagcgag ctgtacaagt acaagtggtt ggcactcag 1440
ccccctgcgc tggcgcacc ccaaggccag cgcgcgctgg tgcagcgca gaagcgccc 1500
gtgggctctg gcgcctgtt catcggttc ctggcgccgc cggggagcac catggcgccc 1560
gectctgta cctcgacct gcagccccc cagcactgta cgggcatcgt gcagcagcag 1620
aacaacctgc tgcgcgccat cgagccccc gacacctgc tgcagctgac cgtgtggggc 1680
atcaagagcg tgcagccgc catcctggcg gtggagcgtt acctgaagga ccagcagctg 1740
ctgggcatct ggggctgcag cggcaagctg alctgcacca ctggaacagc 1800
agctggagca acaagagcct gaccgagatc tgggaacaa tgacctggat ggaagtggag 1860

```

```

cgcgagatcgc gcaactacac cggcctgac tacaacctga tcgagatcgc ccagaaccag 1920
caggagaaga acgagcagga gctgctggag ctggacaagt gggccagcct ctggaactgc 1980
ttcgacatca ccaactggct gtgtgacatc cgcattctca tcatgatcgt gggcgccgtg 2040
atcgcgctcgc gcatcgtgtt cgcctgctgc agcatcgtga accgctgtgc gggcggtcat 2100
agcccccata cctctcgagac cgcctgccc gccccagcgc gcccccagcc ccccgaggcg 2160
atcgaggagg agggccgcga gcgcgacgc gaccgcagca accgcctggg gcacggcctg 2220
ctggccctga tctgggacga cctgcgcagc ctgtgcctgt tcagctacca cgcctcgcgc 2280
gacctcgtgc tgatcgtggc cgcgcatcgt gagctcgtgg gccgcgcggg ctggggagccc 2340
ctgaagtact ggtggaaact gctgcagtae tggagccagg agctgaagag cagcgccgtg 2400
agcctgttca acgccaccgc catcgcctgt gccgagggca cgcacccgat catcgagatc 2460
gtgcagcgca tcttcgcgc cgtgatecac atccccgcc gcacccgcca gggcctggag 2520
cgcgcctcgc tgaagatat cggatcctct aga

```

<210> 67

<211> 2340

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modUS4.delV1/V2

<400> 67

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cgactctctg tttcccccac cgcacccacc gtgctgtggg tgacctgtga ctacggcgtg 120
cccggtggga aggagggcca caccaccctg ttctgcgcca gcgacgccaa ggtctacaag 180
gcgagggccc acaactggtg ggcacccacc gctctgctgc ccaccgaccc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagacaaa catggtggag 300
cagatgcgat aggaatcat cagcctgtgg gaccagagcc tgaagccctg cgtggcgccc 360
ggccaggcct gccccaaggt gagcttcgag cccatcccca tccactactg gcccccgcgc 420
ggcttcgcca tctgaagtg caaggacaag aagttcaacg gcaccggccc ctgcgaagac 480
gtgagcacgg tcagtgcaac ccacggcatc cgcctcgtgg tgagcaccca cgtgctgctg 540
aacgcgcagc tggcgcaggga ggagatcgtg ctgcgctccg agaacttcaac cgacaacgcc 600
aagaccatca tctgtcagct gaacgagtc cgtggagatca actgcatccg ccccaacaac 660
aacacgcgta agaatcatca catcggcccc ggcgcgcctt tctacgccac cgcgcacatc 720
atcggcgaca tccgccaggg ccaactgcaac atcagcaagg ccaactggag caacaccctc 780
cagcagatcg tggagaagct cgcgcgagcag ttccgcaaca acaagacctt catcttcaac 840
agcagcagcg cgcgcgaccc cgcagatcgtg ttccacagct tcaactgcgg cggcgagttc 900
ttctactgca acaccagcca gctgttcaac agcaactgga acatcaccca ggaggtgaac 960
aagaccgaag agaaccgacac catcatactg cctgcgccca tccgcagat catcaacatg 1020
tggcaggagg tggccaaggg catgtacgcc cccccatcc ggcggcaagt caagtgcage 1080
agcaaatatt ccggcctgct gctgacccgc gacggcgcca ccaacaacaa ccgcacacac 1140
gacaccgaga ccttccgccc cgcgcggcgc aacatgaaag acaactggcg cagcgagatg 1200
tacaagtaca aggtggtgcg catcgagccc ctggcgctgg cccccaccca ggccaagcgc 1260
cgcgtggtgc agcgcagaaa gcgcgcctg ggcctggggc tccgtgaccc tgacctgtga ggcccgccag 1320
ggcgccgcgc ggagcaccat gggcgccgcg accctgctgc gcgccatcga ggcccagcac 1380
ctgctgagcg ccatcgtgca gcagcagaac aacctgctgc agccctgctg gcgccagcac 1440
caactcgtgc agctgcacgt gtggggcatc aagcagctgc agggcccgat cctggtcgtg 1500
gagcgtaccc tgaaggacca gcagctcgtg ggcactgggg gctgcagcgc caagctgatc 1560
tgccaccaca cgtgcccctg gaacagcagc tggagcaaca agagcctgac cgcagatctg 1620
gacaaatgca cctggatgga gtgggagcgc gagatcggca actacacccg cctgatctac 1680
aacctgatcg agatcgccca gaaccagcag gagaagaagc agcagagagc gctggagctg 1740
gacaaagtgg ccagcctgtg gaactggctt gacatcacca actggctgtg gtacatccgc 1800
atccttcatc tgatcgtggg cgcctgctgc ggcctgcgca tegtgttcgc cgtgctgagc 1860
atcgtgaaac gcgtgcgcca ggcctacagc cccatcagcc tcgagacccc cctgcccgcg 1920
cagcgcggcc ccgacgcgcc caggggcacg gaggaggagg gcggcgagcg gcacgcgcac 1980
cgcagcaacc cctgtgtgca cgcctgctgc gccctgatct gggagcagct gcgcagcctg 2040
tgctctgtca gctacacacg cctgcgagac ctgctgctga tegtggcccg catcgtggag 2100
ctgctggccc gccgcgctg gagggcctg aagtactggt ggaacctgct gcagctactg 2160

```

```

agccaggagc tgaagagcgc cgcctgagc ctgttcaacg ccaccgccat cgcctgggccc 2220
gaggggcaccg accgcctcat cgagatcgtg cagcgcattc tccgcgcgtg gatccacatc 2280
ccccgccgca tccgccaggc cctggagcgc gccctgctgt aagatatcgg atctctctaga 2340

```

<210> 68

<211> 2385

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modUS4del 128-194

<400> 68

```

gaattcgcca ccatggatgc aatgaagaga gggtctgtgt gtgtgtgtgt gctgtgtgga 60
cgagctcttg ttctgcccag cggcaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
ccggtgtgga aggaggccac caccaccctg ttctgcccga cgcagcccaa ggcttacaag 180
gcgagggccc acaacgtgtg gggcaccacc gcctgctgtg ccaaccgccc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaactgtg ggaagaacaa catggtggag 300
cagatgcgat aggacatcat cagcctgtgtg gaccagagcc tgaagccctg tgaagcgtg 360
accccctgtg cgtgtggggc agggaaactgc gagaccagcg tgatcaccca ggccctgccc 420
aaggtgagct tcgagcccat ccccatccac tactgogccc cgcgcgctt ggcacactg 480
aagtgcgaag acaagaagtt caacggcacc ggccctctga agaactgtg cagcctggcc 540
tgaccccccg gcctccgccc cgtggtgagc acccagctgc tgctgaacg cagcctggcc 600
gaggaggaga tctgtctgog ctccgagaaac ttaccgcaca acgccaagac catcatcgtg 660
cagctgaacg agtctgtgga gatcaactgc atccgcccga acaacaacac gcgtgaagag 720
atccacatcg gccccggcgc cgccttctac gccaccgccc acatcatcgg cgacatccg 780
caggcccact gcaacatcag caaggccaac tggaccaaca ccctcgagca gatctggag 840
aagctcgccg agcagttcgg caacaacaag accatcatct tcaacagcag cagcggcggc 900
gaccccgaga tctgtttcca cagcttcaac tgccggcggc agttcttcta ctgcaacacc 960
agccagctgt tcaacagcac ctggaacatc accgagaggg tgaacaagac caaggagaac 1020
gacacacatca tctgtccctg ccgcatccgc cagatcatca acatgtggca ggaggtgggc 1080
aaggccatgt acgccccccc catccgcggc cagatcaagt gcagcagcaa tattaccgpc 1140
ctgctgtcga cccgcgacgc cggcaccaac aacaaccgca ccaacgacac cgagaccttc 1200
cgcccccggc gccgcaacat gaaggacaac tggcgccagc agctgtacaa gtacaaagtg 1260
gtgcgcatcg agccctggg cgtggccccc acccaggcca agcgccgctg ggtgcagcgc 1320
gagaagcgcg ccgtgggccc gggcgccctg ttcatcggtt tcttggggcg cggcggaagc 1380
accatggcgc cgcctccgt gaacctgacc gtgcaggccc gccagctgct gagcggcatc 1440
gtgcagcagc agaacaaact gctgcgcgcc atcgaggccc agcagcaact cctgcagctg 1500
accgtgtggg gcatcaagca gctgcaggcc cgcactcgtg ccgtggagcg ctacctgaag 1560
gaaccagcag tctgtgggat ctggggcgtc agcgccaaag tgactcgcac caccacgtgt 1620
ctctggaaca gcagctggag caacaagagc ctgaccgaga tctgggacaa catgacctgg 1680
atggagtggg agcgcgagat cggcaactac accggcctga tctacaacct gatcgagtc 1740
gcccagaacc agcaggagaa gaacgagcag gagctgctgt agctggacaa gtggggcagg 1800
gcttggaaact ggttcgacat caaccaactg ctgtggatca tccgcatctt catcatgat 1860
gtggggtggc tgatcgccct gcgcactcgt ttgccctgct tgagcatcgt gaaccgctg 1920
cgccagggct acagcccat cagcctgcag acccgctgc ccgcccagcg cggcccccag 1980
cgccccgagg gcatcgagga ggaggggcgc gagcgcgacc gcgaccgagc caaccgctg 2040
gtgcacggcc tgctggccct gatctgggac gaacctgcga cctgtgctc gtccagctac 2100
cacgcctcgc gcgacctgct gctgatcgtg gcccgcatcg tggagctgct gggccgcgc 2160
ggctggaggc cctggaagta ctgtgggaac ctgctgcagt actggagcca ggagctgaag 2220
agcagcgccg tgagcctgtt caacgcccac gccatcgccc tggccgaggg caccgaccgc 2280
atcatcgaga tctgtgcagc catcttcgcg cgcgtgatcc acatcccccg ccgatccgc 2340
caggsgcctg agcgcgcctt gctgtaagat atcggtatct ctaga 2385

```

<210> 69

<211> 144

<212> DNA

<213> Human immunodeficiency virus

<400> 69
 gacaccatca tectgccctg ccgcatccgc cagatcatca acatgtggca ggaggtgggc 60
 aaggccatgt acgcccccc catccgcggc cagatcaagt gcagcagcaa catcacccgc 120
 ctgctgtctga cccgcgacgg cggc 144

<210> 70

<211> 144

<212> DNA

<213> Human immunodeficiency virus

<400> 70
 ggaactatca cactcccatg cagaataaaa caaattataa acaggtggca ggaagtagga 60
 aaagcaatgt atgccccctc catcagagga caaattagat gctcatcaaa tattacagga 120
 ctgctattaa caagagatgg tggc 144

<210> 71

<211> 144

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic Env
 US4 common region

<400> 71
 gacaccatca tectgccctg ccgcatccgc cagatcatca acatgtggca ggaggtgggc 60
 aaggccatgt acgcccccc catccgcggc cagatcaagt gcagcagcaa catcacccgc 120
 ctgctgtctga cccgcgacgg cggc 144

<210> 72

<211> 144

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic Env
 SF162 common region

<400> 72
 ggcaccatca cctgccctg ccgcatcaag cagatcatca accgctggca ggaggtgggc 60
 aaggccatgt acgcccccc catccgcggc cagatccgct gcagcagcaa catcacccgc 120
 ctgctgtctga cccgcgacgg cggc 144

<210> 73

<211> 4766

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 gp160.modUS4.gag.modSF2

<400> 73
 gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtctgct gctgtgtgga 60
 gcaatcttcg ttctgcccag cgcaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
 ccggtgtgga aggagggcac caccaccctg ttctgcgcca gcagcgccaa ggcttacaag 180
 gccgaggccc acaacgtgtg ggcaccacc gcctgctgtc ccaccgaccc caaccgccag 240
 gaggtgaacc tgaccacgt gaccgagaaac ttcaacatgt ggaagaacaa catggtgag 300
 cagatgeatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cytgaagctg 360

```

acccccctgt gcgtagacct gaactgcacc gacaagctga ccggcagcac caacggcacc 420
aacagcagcca gcggccaccaa cagcaccagc ggcaccagca gcaccagcac caacgtagcc 480
acagcgtggg agaagatgcc cagggggcag atcaagaact gcagcttcaa catcaccacc 540
agcgtggcgg acaaggtgca gaaggagtac agcctgttct acaagctgga ctctggtgcc 600
atcgacaacg acaacgccag ctaccgcctg atcaactgna acaccagcgt gatcaccacc 660
gcctgcccca aggtgagctt cgagcccatc cccatccaat actgcccgcc cgcgggcttc 720
gccatctcga agtgcaagga caagaagttc aacggcaccg gccctcgcaa caacgtgagc 780
accgtgcagt gccaccaacg catccgcccc gtggtgagca cccagctgct gctgaacggc 840
agcctggcgg agggaggagt cgtgctgcgc tcgagaaact tcacgacaa cgcgaagacc 900
atcatcgtgc agctgaagca gtccgtggag atcaactgca tcgcccccaa caacaacagc 960
cgtaagagca tccaacatcg ccccgccgcg gccctctacg ccaccggcga catcatcgcc 1020
gacatccgcc agggccactg caacatcagc aaggccaact ggaccaacac cctcgagcag 1080
atcgtggaga agctgcgcga cgagtccggc aacaacaaga catcatctt caacagcagc 1140
agcggcggcg accccagagt cgtgtccac agcttcaact gcggcgcgga gttctctac 1200
tgcaacacca gccagctgtt caacagcacc tggaaactca ccgaggaggt gaacaagacc 1260
aaggagaacg accacatcat cctgcctctc cgcctccgcc agatcatcaa catgtggcag 1320
gaggtgagca aggccatgta cgcgcccccc agatcaagtg cagcagcaat 1380
ataccggccc tgctgctgac ccgcgacggc ggcaccaaca acaaccgac caacgacacc 1440
gagaccttcc gcgccggcgg cggcaacatg aaggacaact ggcgcagcga gctgtacaag 1500
tacaaggttg tgccatcga gccctggggc gtggccccca cccaggccaa gcgcgcgctg 1560
gtgcagcgcg agaagcgccg cgtgggcctg ggcgcctgt tcatcgctt cctggggccc 1620
gcggggagca ccatggggcg cgcctccgtg accctgaacg tgacggcccg ccaagctgct 1680
agcgccatcg tgcaagcaga gaacaaacct ctgcgcgcca tcgaggccca gcagcactgt 1740
ctgcagcaga cgcgtggggg catcaagcag ctgcaggccc gcatcctggc cgtggagcgc 1800
tacctgaagg accagcagct gctgggcctc tggggctgga cgcggcaagt gatctgacc 1860
accacgtcgc cctggaaacg cagctggagc aacaagaagc tgaccagat ctgggaacca 1920
atgacctgga ttgagtgga gcgcgagatc ggcaactaca cgcgcctgat ctacaacctg 1980
atcgagatcg cccagaaaca gcaggagaag aacgagcagg agctgtgga gctggacaag 2040
tggggcagcc tggtggaact gtctgacatc accaactggc tbtgtacat ccgcatcttc 2100
atcatgatcg tggcgccgct gatcggcctg cgcactgtgt tcgctgctgt gagcatctgt 2160
aacgcgtgca gccagggcta cagcccccat agcctcgaga cccgcctgcc cgcgccagcg 2220
ggccccgacc ccccgagggg catcgaggag gaggcgcgcg agcgcgaccg cgaaccgacc 2280
aaccccgctg tgcaaggcct gctggccctg atctgggagc acctgcgac cctgtgctgt 2340
ttcagctacc accgctcgcg cgacctgctg ctgactgtgg cccgcactgt ggagctgctg 2400
ggcccgccgg gctgggaggg cctgaaagta ttggtggaacc tgctgagta ctgagaccag 2460
gagctgaaga gcagcgccgt gagcctgttc aaagccaccg ccatcgccgt ggcgagggcg 2520
accgaccgca tcatcgagat cgtgcagcgc atcttccgcg cctgatcca catcccgcgc 2580
cgcatccccc agggccttga gcgcgcctg ctgtaagata tgggactctc tagagaattc 2640
cgcccccccc ccccccccc gcttccctcc ccccccccta atgttatttt ccacatattt gcgctctttt 2700
ttggaataag ccgggtgtgc gtttgtctat atgttatttt ccacatattt gcgctctttt 2760
ggcaattgtga gggcccgga accctggcct cgtctcttga cgagcatctc taggggtttt 2820
tccccctcgg ccaaggaaat gcaaggctgt ttgaatgtcg tgaaggaaag agtctctctt 2880
gaagcttctt gaagacaacac aacgtctgta gcgacctctt cgaggcagcg gaacctctc 2940
cctggcagca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc tgcgaaggcg 3000
cccgcccgca agtgcacagt tbtgagttgg atagtttggg aaagagtcaa atcctctctc 3060
tcaagcgtat tcaacaaggg gctgaaggat gccccaaggg taccaccttg tatgggactc 3120
gatctggggc ctcggtgcac atgctttaca tbtgtttagt cgaggttaaa aaaactctta 3180
ggccccccga acccggggga cgtggttttc ctttgaaaaa cagcataata ccatggggcg 3240
cgcgcacag gtgctgagcg cggcgagcgt ggacaagtgg gagaagatcc gctcgccccc 3300
cggcgccagc aagaagtaca agctgaagca catcgtgtgg gccagccgcg agctggagcg 3360
cttcgcogtg aacccccggc tctggagag cagcgagggg tcgcccgaga tcttgagcca 3420
gctgcagccc agcctgcaga cggcgacgga ggagctgcgc agcctgtaca acaaccgtgc 3480
cacctcgtac tctgtgcacc agcgcatcga cgtcaaggac accaaggagg cccctgagaa 3540
gatcgaggag gacgagaaca agtccaagaa gaaggccacg cagggcgccg ccgcccgcgg 3600
cacccggcaac agcagccagg tgagccagaa ctaccacct gtgcagaacc tgcagggcca 3660
gatggtgcac caggccatca gcccccgcac cctgaagcgc ctgagccgag tgggtgaagg 3720
gaaggccttc agccccgagg tgatccccat gtcagccgc ctgagccagg gcgcccaccc 3780
ccaggacctg aacacagatg tgaacacctg gggcgccac caggccacca tgcagatgt 3840
gaaggagacc atcaacgagg aggcggccga gtgggagcgc tgcaaccccg tgcagccggg 3900

```

```

ccccatcgcc ccggcccgaga tgcgcgagcc ccgcggcagc gacatcgccg gcaccaccag 3960
accctcgag gagcagatcg gctggatgac caacaacccc cccatccccg tggcgagagat 4020
ctacacggcg tggatcatcc tgggcctgaa caagatcgct cggatgtaca gccccaccag 4080
catcctggac atccggccagg gccccaagga gcccttcggc gactacgttg accgctttcta 4140
caagaccctg ccgctgagc aggcagccca ggacgtgaag aactggatga ccgagacctt 4200
gctgggtgag aacgccaacc ccgactgcaa gaccatcctg aaggcttctg gcccccgggc 4260
cacctgggag gagatgatga ccgcttgcca gggcgtgggc ggcccgggcc acaaggcccg 4320
cgtgctggcc gaggcgatga gccagggtgac gaaccggcgcc accatcatga tgcagcgccg 4380
caacttcgcg aaccgcgga agacgctcaa gtgcttcaac tgcggcaagg aggggccacat 4440
gcggcgaagc tgcgcgccc ccgcgaagaa gggctgctgg cgctgcgccc gcgagggcca 4500
ccagatgaag gactgcaccg agcgccaggc caacttctct gcgaagatct gccccagcta 4560
caaggcgccg ccggcgaact tctcgagag ccgcccgcag cccaccgccc cccccgagga 4620
gagcttcgcg ttccgcgagg agaagaccac ccccgccag aagcaggagc ccatcgacaa 4680
ggagctgtac cccctgacca gctcgccgag cctgttcggc aagcacccca gcagccagta 4740
agaattcaga ctcgagcaag tctaga 4766

```

<210> 74

<211> 4689

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modSF162.gag.modSF2

<400> 74

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
cagctcttcg ttctgcccag ccgctgtggg aagctgtggg tgaccgtgta ctacggcgctg 120
ccgctgtgga aggagggcac caccaccctg ttctgcccga gcgacggccaa ggctctacgac 180
accgaggtgc acaacgtgtg ggcaccacca gctcggtgct ccaccgaccc caaccgccag 240
gagatcgctg ttgagaacgt gaccgagaga ttcaacatgt tgaagaacaa catgtgtggag 300
cagatgcacg aggcacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaaagctg 360
accctccctg cgtgtaccct gcaactgcac aacctgaaga acgcccacaa caccagagac 420
agcaactgga aggagatgga ccgcgccgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtag gccctgttct acaagctgga cgtgtgtccc 540
atcgacaacg acaaccaccg ctacaagctg atcaactgca accaccagct gatccccac 600
gcttgcceca aggtgagctt cgagcccatc ccatccact actgogcccc gccgggcttc 660
gcatcctgca agtgcaacga caagaagttc aacggcagcg gccctgcac caacgtgagc 720
accgttcctg gcaccacgg catccgcccc gtggtgagca cccagctgct gctgaaacggc 780
agcctggcgc aggagggcgt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc 840
atcatctgac agctgaagga gagcgtggag atcaactgca ccgcccacaa caacaacacc 900
cgcgaagaga tcaccatcgg ccccgccgcg gcctttctac ccaccggcga catcatcagg 960
gacatccgac agcccccact caacatcagc ggcgagaagt ggaacaacac cctgaagcgc 1020
atcgtgacca agctgcagcg ccagttcggc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcgcgaccc ccgagatcgt gatgcacagc ttcaactgcy gcggcgagtt ctctactgc 1140
aaccagaccg agctgttcaa cagcacctgg aacaacacca tcggccccaa caacaccac 1200
ggcacataca cctggccctg ccgcatcaag cagatcatca accgtggca cctgaagcgc 1260
aaggcagatg acgccccccc catccgcgcg cagatccgct gcagcagcaa catcacggcg 1320
ctgctgctga ccccgacagg cggcaaggag atcagcaaca ccaccagatg ttctccgccc 1380
ggcgcgccgc acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtgtgtgag 1440
atcgagcccc tgggtgtgcy cccaccaaag gccaaagccc tgggtgtgca gccgagaaag 1500
cgcgccgtga cctggggcgc catgttctct ggccttcttg gcgcgcggcg cagcaccaatg 1560
ggcgccccga gctgacccct gacgctgcag gcccgccgca tgcctgagcg catctgtcag 1620
cagcagaaca acctgctgcy gcccatcgag gcccgagcgc acctgctgca gctgaccgtg 1680
tggggcatca agcagctgca ggcggcgctg ctggcgctgg agcgctacct gaagagaccg 1740
cagctctgcy gcatctgggg ctgcagcgcg aagctgatct gcaccacgcg cgtgcctcgt 1800
aaccggcagct ggagcaacaa gagcctggag cagatctgga acaacatgac ctggatggag 1860
tgggagcgcg agatcgacaa ctacaccaac ctgacttaca cctgatcga ggaagaccag 1920
aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980

```

```

aactgggtcgc acatcagcaa gtggctgtgg tacatcaaga tcttcatcat gatcgtgggc 2040
ggcctgggtgg gctcgcgcac cgtgttcacc gtgctgagca tctgtaaccg cgtgcgcagc 2100
ggctacagcc cctcagcctt ccagaccgcc ttccccgcc cccgcggccc cgtgtgacac 2160
gagggcagct aggaggaggg cggcgagcgc gaaccgcgac gcagcagccc cctgtgtgac 2220
ggcctcgtcgc cctcagcctg ggaagacacg cgcagcctgt gctgttcag ctaccaccgc 2280
ctcgccgacc tgatcctgat cgcgcgccgc atcgtggaag tgctgggccc cgcgcgctgg 2340
gagggccctga agtactgggg caacctcgtc cagtaactga tccaggagct gaagaacacg 2400
ccgctgagcc tgttcagacc catcgccatc gccctggccc agggcaccga ccgcacatc 2460
gaggtgggccc agcgcctcgg ccgcgccttc ctgcacat cc ccgcgcgcat ccgcacaggc 2520
ttcgagcgcg cctcgtctga actcgagcaa gtctagagaa ttccgccccc cccccccccc 2580
ccccctcccc tcccccccc ctaacgttac tggccgaagc cgcttggaa taaggccgct 2640
tgcttttctc tatatgttat ttccacacat attgcgtctc ttggcaaatg tgagggcccg 2700
gaaacctgct cctgtctctc tgacgagcat tctcagggtt ctttccccct tcgccaaagg 2760
aatgcaaggt ctgttgaagt tcgtgaagaa agcagttcct ctggaagctt cttgaagaca 2820
aacaacgtct gtacgcaccc ttgcaggcca cgggaacccc ccacctggcg acaggtgcct 2880
ctcgccgcaa aagccacgtg tataagatac acctgcaaa ggcggcacaac cccagtgcga 2940
cgttgtgagt tgcatatgtt tggaaagagt caaatggctc tctcaagcgc taattcaaca 3000
ggggctggaag gatgcccaaa aggtacccca ttgatggga tctgatctgg ggctcgggtg 3060
cacatgcttt acatgtgttt agtcgaggtt aaaaaaacgt ctaggccccc cgaaccacga 3120
ggcagctggtt ttctttttaa aaacacgata ataccatgg cgcccgccc agctgtcgtg 3180
ggcgccgcgc gctggacaag tgggagaaga tccgcctgcg ccccgccgpc aagaagaagt 3240
acaagctgaa gcacatctg tggccagcgc gcgagctgga gcgcttcgcc gtgaaccccg 3300
gactgctgga gaccagcgag ggctgcgcgc agatcctggg ccagctgcag cccagctgcg 3360
agccggcgag cgaggagctg cgcgcgctgt acaacacgtt ggcacacctg tactgcgtgc 3420
accagcgcat cgagctcaag gacaccaagg aggccttga gaagatcgag gaggagcaga 3480
accaagtcaa gaagaaggcc cagcaggccc ccgcgcgcgc cggcaccgag aacagcagca 3540
aggtgagcca gaactacccc atcgtgcaga acctgcaggg ccagatggtg caccagcca 3600
tcagcccccgc caccctgaac gcctgggtga aggtggtgga ggagaaggcc ttacgccccg 3660
aggtgatccc catgttcacg gccctgagcg agggcgccac cccccaggag ctgaacacga 3720
tggtgaacac cgtgggcggc caccaggccc ccatcgagat gctgaaggag accatcaag 3780
aggaggcgcg cagtgaggac cgcgtgcacc cgtgacacg cggccccctc gccccggggc 3840
agatgcgcga gcccccggcg agcgacatc cgcgcaccac cagcaccctg caggagcaga 3900
tcggtcgtgt gaccaacaac ccccccatc cctgtgggca gatctacaag cgtggtatca 3960
tctcgggcct gaacaagatc gtgcgagatg acagccccac cagcatcctg gacatccgcg 4020
agggcccca aaaggccctc cgcgaactac tggaccgctt ctacaagacc ctgcgcgctg 4080
agcaggccag ccaggacgtg aaagaactga tgaccagag cctgctggtg cagaacgcca 4140
accccagctc caagaccatc ctgaaggctc tcggccccgc ggcacacctg gagagagata 4200
tgaccgcctg caagggcgtg ggcggccccc gccacaaggc ccgctgtgct gccagggcga 4260
tgagccaggt ggaagaacct gcgaccatca tgatgcagcg cggcaacttc cgcaaccagc 4320
ggaagcagct caagtcttc aactgcggca aggaaggcca caccgcaggc aactgcgcgt 4380
ccccccgcaa gaagggctgc tggcgctgcg gccgcagggg ccaccagatg aaagactgca 4440
ccgagcgcca ggccaacttc ctggggcaaga tctggccagc ctacaaggcg cgcgccggca 4500
actctctgca gggccgcccc gagccccacc cccccccgca ggagagcttc cgtctggggc 4560
aggagaagac cacccccagc cagaagcagg agccccctga caaggagctg taccocctga 4620
ccagctcgcg cagcctgttc ggcaacgacc ccagcagcca gtaaggaaatc agactcgagc 4680
aagtctaga

```

<210> 75

<211> 4472

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modUS4.delV1/V2.gag.modSF2

<400> 75

```

gaattccgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgctg gctgtgtgga 60
gcagctcttgc ttctgccagc cgcaccaccc gtgctgtggg tgacctgtga ctacggcggtg 120

```


ccggtgtgga aggaaggccac caccaccctg tcttgcgcca gcgaecgeaa ggcctacaag 180
 gccagggccc acaacgtgtg ggccacccac ggcctgcgtgc ccacgcgccc caacccccag 240
 gaggtgaacc tgaccaacgt gaccgagaaac ttcaacatgt ggaagaaaca catgtgtgag 300
 cagatgcgat aggaacatcat cagcctgtgg gaccagagcg tgaagccctg cgtgtggcgc 360
 ggccaggcct gccccaaggt gagcttcgag cccatcccca gcaactactg gcccccgcgc 420
 ggcttcgcca tccctgaagt caaggacaag aagttcaacg gcaacggccc ctgcgaagac 480
 gtgagcaccg tgcagtgcac ccacggcatc cgcgccgtgg tgagaaccca gctgtgtctg 540
 aacggcagcc tggcgagga ggaagatcgt ctgcgctcgc agaactcac cgacaagccc 600
 aagaccatca tgtgtcagct gaacgagtc ccgtggagatc actgcatcgc ccccaacaac 660
 agacgcgcta agagcatcca catcgccccc ggccgcgcct tctacgcacc cggcgacatc 720
 atcgccgata tcccgcaagg ccactgcacac actcagcaag ccaactggac caacaccctc 780
 gagcagatcg tggagaagct gcgcgagcag ttccggcaaca acaagaacct catcttcaac 840
 agcagcagcg gcggcgaccc cgagatcgtg ttccacagct tcaactgcgc cggcgagttc 900
 ttctactgca acaccagcca gctgttcaac agcacctgga acatcaccca ggaggtgaac 960
 aagaccaagg gaaacgcacac catcatcctg cccctgcgca tccgccaagt catcaacatg 1020
 tggcaggagg tgggcaaggc catgtacgcg ccccccatac gcggccagat caagtgcagc 1080
 agcaatatta ccggcctgct gctgaccgcg gacggcgcca ccaacaacaa ccgaccacac 1140
 gacaccgaga ccttccgccc cggcgccggcg aacatgaagg acaactggcg cagcgagctg 1200
 tacaagtaca agtgtgtgct catcgagccc ctggcgctgg ccccccacca ggccaagcgc 1260
 cgcgttgtgc agcgcgagaa gcgcgcgcgt ggccctgggg cccgtgttcat cggcttctgt 1320
 ggccgcgcgc ggaacacat ggccgcgcgc ggccgcgcgc tgaccgtgca ggccgcgcgc 1380
 ctgctgagcg gcatcgtgca gcagcagaac aaactcgtgc gcgccatcga ccccgacag 1440
 caactcgtgc agctgacgtg gtggggcctc aagcagctgc aggcgcgcct cctggcgcgt 1500
 gagcgttacc tggagacca gcagctcgtg ggcacttggg gctgcagcgc gcaagctgatc 1560
 tgcaccacca cctgcccctg gaacagcagc tggagcaaca agagcctgac cgagatctgg 1620
 gacaacatga cctggatgga gtgggagcgc gagatcggca actacacgcg cctgatctac 1680
 aacctgatcg agatgcacca gaaccagcag gaagagaagc agcaggagct gctggagctg 1740
 gacaagtggg ccagcctgtg gaactggctc gacatcacca actggtgtg ctgacatcgc 1800
 atcttcatca tgatcgtggg cggcctgatc ggcccgcca tctgttgcgc cgtgctgagc 1860
 atcgtgaacc cgcgtgcgca gggtcacagc cccatcagcc gtcagacccc cctgccgcgc 1920
 cagcgcgcgc ccgaccgcc ccgagggcatc gaggaaggag gcgcgcagcg cgaccgcgac 1980
 cgcaacaacc gcttgggtgca cggcctgctg gccctgatct gggaagacct gcgcagcctg 2040
 tgcctgttca gctaccacgc cctgcgcgac ctgctgctga tctgtggccc catcgtggag 2100
 ctgctggggc gccgcggctg ggagggccctg aagtaactgtt ggaacctgct gcaactctgg 2160
 agccaggagc tgaagagcag cgcctgagc ctgttcaacg ccaaccgcat ccgcgtggcc 2220
 gagggcacgc accgcacatc cgagatcgtg cagcgcatct tccgcgcgct gaaccacatc 2280
 ccccgccgca tccggcaggc cctggagcgc gccctgctgt aagatatcgg atctctgga 2340
 gaattccgcc ccccccctct cccctccccc cccctaacgt tactggcgca 2400
 agccctgttg aataaggcgc gtgtgcgttt gtctatatgt tattttccac catattgccc 2460
 tctttgttga atgtgagggc cgggaaaact ggccctgtct tcttgacgag catctctagg 2520
 ggtctttccc ctctgcgcaa aggaatgcac ggtctgttga atgtgtgaa ggaagcagtt 2580
 cctctggaaag ctcttgtgaag acaaaacacg tctgtagcga ccttttgacg gcagcggaac 2640
 cccccacgc gcgacaggtg cctctgcggc caaaagccac gtgtataaga tacacctgca 2700
 aaggcgcaac aaccccagtg ccacgcttgt agttggatag ttgtgaaagc agtcaaatgt 2760
 cctctctcaa gcgtattcaa caaggggctg aaggatgccc agaagttacc ccattgtatg 2820
 ggatctgac tggggccctg gtgcacatgc tttaactgtg tttagtgcag gttaaaaaaa 2880
 cgtctaggcc ccccgaaacca cggggacgtg gttttcctt gaiaaaacacg caattgtatg 2940
 ggccgcgcgc gccagcgtgc tgagcgccgg cgagctggac aagtgggaga agatccgcct 3000
 ggcgcgcgcg ggcaagaaga agtacaagct gaagcacatc gtgtgggcca gccgcagcgt 3060
 ggagcgtctc gcgctgaacc cggcctgct ggagaccagc gaagggctgcc gccagatcct 3120
 ggccagagctg cagccagcgc tgcagaccgg accgagagc ctgcgcagcc tgtacaacac 3180
 cgtgcccacc ctgtactgcg tgaccacgcg catcgacgtc aaggaacaca aggaaggcct 3240
 ggagaagatc gaggaagagc agaacaagtc caagaagaa gccacgcagg ccgcgcgcgc 3300
 gcccgccacc ggcaacagca gccaggtgag ccagaactac ccaatcgtgc agaacctgca 3360
 ggccagatg gtgcaccagg catcagccc ccgcacccctg aacgcctggg tgaaggtgtg 3420
 ggagagaaag cctctagacc ccgaggtgat ccccatgttc agcgcctgga agcggggcgc 3480
 cacccccag gacctgaaca cgatgttgaa caccgtgggc ggccaccagg gcccgatgca 3540
 gatgtgaa gaaacatca acgaggagcc gcccgagtgg gaccgctgc acccctgca 3600
 cgcggcgccc atcgccccgc gccagatcgc cgagccccgc ggcagcgaca tgcgcggcac 3660

```

caccagcacc ctgcaggagc agatcggtct gatgaccaac aaccccccca tccccgtggg 3720
cgagatctac aaggcgttga tcatcctggg cctgaacaag atcgtgcgga tgtacagccc 3780
caccagcacc ctgcagatcc gccaggggccc caaggagccc ttccgcgact acgtggagcc 3840
ctttacaaag acctgcgcgc ctgagcaggc cagccaggac gtgaagaact ggtgagccga 3900
gacctctgct gtgcagaacg ccaaccccga ctgcaagacc atcctgaagg ctctcgccccc 3960
cgcgccgccc ctgcaggaga tgatgaccgc ctgccagggc gtggggcgccc ccggccacaa 4020
ggcccgcgtg ctggccgagg cgtatgagca ggtgacgaac ccggcgacca tcatgatgca 4080
gcgcgggcaac ttccgcaccc agcgggaagac cgtcaagtgc ttcaactcgc gcaaggaggg 4140
ccacaccgcc aggaactgcc gcgccccccc caagaagggc tctgtgcgct gcggccgcca 4200
gggcaccacag atgaaggact gcaccgagcg ccaggccaac ttctgggca agatctggcc 4260
cagctacaag aggcgcggcc gcaacttctc gcagagcgcc ccgagcccca ccgccccccc 4320
cgaggagagc ttccgcttgc gcgaggagaa gaccaccccc agccagaagc agggagccat 4380
cgacaaggag ctgtaccccc tgaccagcct gcgcgacct ttgcggcaag accccagcag 4440
ccagtaagaa ttcagactcg agcaagtcta ga

```

<210> 76

<211> 4608

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modSF162.delV2.gag.modSF2

<400> 76

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtctgct gctgtgtgga 60
gcagctcttc ttccgcccgc cgccgtggag aagctgtggg tgaccgtgta ctacggcgctg 120
cccggtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgccc aaagcctacgac 180
accgaggtgc acaacgtgtg ggccacccac gctgtgctgc ccacgagccc caacccccag 240
gagatcgtgc tggagaacct gaaccgagaa ttcaactgtg ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accgccctgt gctgtaccct gcactgcacc aacctgaaga acgcccacaa cccaagagac 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtggggccc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccagg cctgccccaa ggtgagcttc 540
gagccccatc ccatccacta ctgcgcccc gcgcgcttgc ccatcctgaa gtgcaacgac 600
aagaagtcca acggcagcgg cccctgcacc aactgagca cgtgacagc gctgcagctg 660
atccgccccg tggtagacac ccagctgctg ctgaacggca gcctggccga ggaggggcgtg 720
gtgatccgca gcgagaactt caccgacaac gccaaagcca tcatcgtgca gctgaaggag 780
agcgtgggca tcaactgcac ccgcccacac acaaacaccc gcaagagcat caccatcgcc 840
cccgcccgcg ccttctacgc caccggcgac atcatcgggc acatccgcca ggcccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tctgtacca gctgcaggcc 960
cagttcgcca acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcaccca cgtgttcaac 1080
agcacttgga acaacaccat cggcccacac aacaccaagc gcacatcac cctgcccctg 1140
cgcatcaagc agatcataca ccgctggcag gaggtgggca agggcatgta cggccccccc 1200
atccgcgccc agatccgctg cagcagcaac atcaccggcc tctgctgac ccgagcagcg 1260
ggcaaggaga tcaacaacac caccgagatc ttccgcccgc cggcgcgga catcgcgca 1320
aactggcgca gcgagctgta caagtacaag gtggtgaaga tgcagcccc ttggcgtgccc 1380
cccaccaagg ccaagcgccc cgtggtgcag cgcgagaagc gcgcctgag cctggcgcc 1440
atgttctcgt gcttctctgg gcgcgcccgc agcaccatgg gcgcccgcag cctgacctgt 1500
accgctcagg ccgcgcagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
gccatcgagg ccagcagcga cctgctgcag ctgacctgtg ggggcatcaa gcagctgcag 1620
gcccgcgctc tggccgtgga gcgtacctg aaggaccagc agctcgtggg catctggggc 1680
tgacggcgca agctgatctg caaccaccgc gtgccctgga acgcaagctg gaagcaaca 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgga gatcgaca 1800
tacaaccaac tgatctacac cctgatcgag cagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtggccc agcctgtgga actggttcca catcagcaa 1920
tggctgtggt acatcaagat cttcatcatg atcgtggggc gctggtggg cctggcgcat 1980
gtgttcacgc tctgtgagcat cgtgaaccgc gtcgcccagg gctacagccc cctgagcttc 2040

```

```

cagaccgcgt tccccgcccc ccgcggcccc gaccgccccg agggcatcga ggagagaggg 2100
ggcgagcgcg acccgagccg cagcagcccc ctggtgcaag gctctgtggc cctgactctg 2160
gacgacctgc gcagcctgtg cctgttcagc taccaccgce tgcgcgacct gatcctgatac 2220
gcgcgccgca tcgtggagct gctggggcgc cgcggtctgg agggcctgaa gtactggggc 2280
aacctgctgc agtactggat ccaggagctg ccagacagcg ccgtgagctg gtcgcagcgc 2340
atcgccatcg ccgtggccga gggcagccag cgcctatcag aggtggccca gcgcagcgcg 2400
cgcgctcttc tgcacatccc ccgcgcgcat cgcgcagggt tcgagcgcgc cctgctgtaa 2460
ctcgagcaag tctagagaa tccgcgcccc cccccccccc cccctctccc cccccccccc 2520
taacgtttact ggccgaagcc gcttggaaat agggcggtgt gcggttgtct atatgttatt 2580
ttccaccata ttgcgctctt ttggcaatgt gaggggccgg aaacctgggc ctgctctctt 2640
gacgagcatt cctaggggtc ttccccctct cgccaaagga atgcaaggtc ttgttgaaat 2700
cgtgaaggaa cgagttcttc tggaaagctt ttgaagacaa acaacgtctg tagcgacct 2760
ttgcaggcag cggaaccccc cactggcgga caggtgcctc tgcggccaaa agccacgtgt 2820
ataagataca cctgcaaaag ggccacaacc ccagtgcacc gttgtgagtt ggatagttgt 2880
ggaaagagtc aaatggctct cctcaagcgt attcaacaag gggctgaagg atgcccagaa 2940
ggtaacccat tgtatgggat ctgatctggg gctcgggtgc acatgcttta catgtgttta 3000
gtcgagggtta aaaaaacgtc taggcccccc gaaccacggg gacgtggttt tcccttgaaa 3060
aacacgataa taccatgggc gcccgcccca gctgtctgag cggcgggcag cttggacaagt 3120
gggagaagat ccgcctgcgc cccggcgcca agaagaagta caagctgaag cacatcgtgt 3180
ggggcagccg cgagctggag cgcttgccgc tgaacccccg cctgtggag accagcagag 3240
gctgccgcca gatcctgggc cagctgcagc ccagcgctgc gactgcaagg 3300
gcgacctgta caaacacctg gccaccctgt gccaccctgt gacgtgcaag 3360
acaccaaaga ggccctggag aagatcgagg agggacagaa caagtccaag aagaagggccc 3420
agcaggccgc ccgcgcgcgc ggccacggca acagcagcca ggtgagccag acctacccca 3480
ctgctgagaa cctgcaggcc cagatggtgc accaggcata cagccccccc accctgaaag 3540
cctgggtgaa ttgtggtggg gagaaggcct tcagccccga ggtgatcccc atgttcaagg 3600
ccctgagcga gggcgccacc ccccaggacc tgaacacgat gttgaacacc gtggcgggcc 3660
accaggccgc catgacagat ctgaaggaga ccatcaacga ggaaggccgc atgtgggacc 3720
cgctgcaccc cgtgcacgcc ggccccatcg ccccccgcca gatgcccagc ccccgcgcca 3780
cgcatccacc ggccacaccc agcacccctg aggagcagat cggctggatg accaacaacc 3840
ccccatccc cgtggcgag atctacaagg gttggtatcat cctggcgctg aacaagatcg 3900
tgcggatgta cagccccacc agcatcctgg acatccgcca gggccccaag gagcccttcc 3960
cgcatcagct ggaccgcttc tacaagacc tgcgcgctga gcaggccagc caggacgtga 4020
agaactggat gaccgagacc ctgctgggtgc agaaccgcaa cccgcactgc aagacacacc 4080
tgaaggctct cggccccgcg gccaccctgt aggaagatgat gaccgctgac caggcgctgg 4140
gcggccccgg ccacaaggcc cgctgtgctg ccgagcgcat gaccaggtg agcacccggc 4200
cgacatcat gatcgagcgc ggcaacttcc gcaaccagcg gaagacgctc aagtgtctta 4260
actcgcccaa ggaaggccac accgccagga actgcccgcg ccccgcaagg aaggcgctgt 4320
ggcgctgcgg ccgcgagggc caccagatga aggaactgcac cgagcgccag gccaaacttcc 4380
tgggcaagat ctggcccagc tacaagggcc gccccggcaa ctctctgcag agcgccccgc 4440
agccccagc ccccccgag gagagcttcc gcttcggcga ggagaagacc acccccagcc 4500
agaagcagga gcccatcgac aaggagctgt acccctgcag cagcctgcgc agccctgttc 4560
gcaacgaccc cagcagccag taagaattca gactcgagca agtctaga 4600

```

<210> 77

<211> 1680

<212> DNA

<213> Human immunodeficiency virus

<400> 77

```

ccattagtgc ctattgaaac tgtaccagta aaattaaagc caggaatgga tggccccaaa 60
gttaagcaat ggccattgac agaagaaaaa ataaaagcat tagtagagat atgtacagaa 120
atggaaaaag aggggaaaaa ttcaaaaatt gggcctgaaa atccatacaa tactccagta 180
tttgcrtata agaaaaaaga cagtactaaa tggagaaaaa tagtagattt cagagaactt 240
aataaaagaa ctcaagactt ctgggaagtt cagtttagaa taccacaccc cgcagggtta 300
aaaaagaaaa aatcagtaac agtatgggat gtgggtgat catactttc agtccctta 360
gataaagact ttgaaagata tactgcattt accataccta gtataaaca tgagacacca 420
gggattagat atcagtaaaa tgtgtgcaca cagggatgga aaggatcacc agccaatttc 480
caaagtatga tgacaaaaat cttagagact tttagaaaaa agaattccga catagtattc 540

```

```

tatcaataca tggatgattt gtatgtagga tctgacttag aaatagggca gcataagaaca 600
aaaatagagg aactgagaca gcactctgtt agtgggggat ttaccacacc agacaaaaaa 660
catcagaaag aacctccatt cctttgggat ggttatgaac tccattctcg taataggaca 720
gtacagccca taatgtctgc agaaaaagac agctgggact tcaatgacat acagaagcta 780
gtgggaaaaa tgaattgggc aagtcagatt tatgcaggga ttaaagtaaa gcagttatgt 840
aaactccctta gaggaaccaa agcactaaca gaagtaatac cactaacaga agaagcagag 900
ctagaactgg cagaaaacag ggagattcta aagaaccag tacatgaagc atattatgac 960
ccatcaaaaag acttagtagc agaataacag aagcaggggc aaggccaatg gacatatcca 1020
atttatcaag agcoatttaa aactctgaaa acaggaagat atgcaaggat aggggtgtcc 1080
cacactaatg atgaaaaaca gttaacagag gcagctgaaa aagatattcc agaaagcata 1140
gtaattatgg gaagatttcc taaatttaaa ctaccatac aaaaaggaaac atgggaagca 1200
gggtgatgg agtattggca agctacctgt attctctagt gggagtttgt caataccctt 1260
cccttagtga aattatggta ccagtttagg aaagaacca tagtaggagc agaaactttc 1320
tatgtatgat gggcagctaa tagggagact aaattaggaa aagcaggata tggttactgac 1380
agaggaagac aaaaagttgt ctccatagct gacacaacaa atcagaagac tgaattacaa 1440
gcattatcct tagctttgca ggattcggga ttagaagtaa acatagtaac agactaccaa 1500
tatgcattag tagcaattca agcacaacca gataaagatt aatcagagtt agtcagtgaa 1560
ataatagagc agttaataaa aaaggaaaaa gtctacctgg catgggtacc agcacacaaa 1620
ggaattggag gaaatgaaca agtagataaa ttagtcastg ctggatcagg gaaagtacta 1680

```

<210> 78

<211> 1865

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: GPI

<400> 78

```

gtcgacgcga ccatggggcgc ccgcccagc gtgctgagcg gcggcgagct ggacaagtgg 60
gagaagatcc ccctgcgcgcc cggcggcga aagaagtaca agctgaagca catcgtgtgg 120
gccacgccgc agctggagcg cttcgccgtg aaccccggcc tgcctggagc cagcaggggc 180
tgccgcgcga tccctggcca gctgcagccc agcctgcaga ccggcagcga ggagctgcgc 240
agcctgtaca acacccgtggc caccctgtac tgcgtgcacc agcgcacgca gctcaaggaa 300
accaagggag ccctggagaa gctcagaggag gagcagaaca agtccaagaa gaaggcccag 360
caggccgcgc ccgcgcgcgc caccggcaac agcagccagg tgagccagaa ctaccccatc 420
gtcgagaacc tgcagggccca gatggtgcac caggccatca gcgcccgcac cctgaacgcc 480
tgggtgaagg tgggtgagga gaaggccttc agccccgagg tgatccccat gttcagcgcc 540
ctgagcgagg gcgcacccc ccaggacctg aacacgatgt tgaacacggt gggcgccac 600
caggccgcga tgcagatgct gaaggagacc atcaacgagg agggcccgca gtgggacgcc 660
gtcacctccc tgcacgcggc cccatcgccc ccggcccgca tgcgcgagcc ccgcgagcgc 720
gcacatcgcc gcacccacag caccctgcag gagcagatcg gctggatgac caacaccccc 780
cccatccccg tggcgagagt ctacaacggg tggatcatcc tgggcctgaa caagatcgct 840
cgcatgtaca cccccacag catcctggac atccgcgagg gccccaagga gcccttcgcg 900
gactactctg accgcttcta caagaccctg cgcgctgagc agggccagca ggaactgaag 960
aactggatga ccgagacctt gctggtgcag aacgccaacc ccgactgcaa gaccattcgt 1020
aaagctctcg gcccccgggc caccctggag gagatgatga ccgctgcca gggcgaggcg 1080
ggccccggcc acaaggcccg cgtgctggcc gaggggatga gccaggtgac gaacccggcg 1140
accatcatga tgcagcggcg caacttcgcg aaccagcgga agacgctcaa ggtcttcaac 1200
tgcggcaagg agggccacac cgccaggaaac tgcgcgccc ccgcgaagaa gggctgctgg 1260
cgctcgccgc gcgaaggaca ccaaatgaaa gattgcactg agagacaggc taatttttta 1320
gggaagattct ggcttctcta caagggaagg ccagggaatt ttcttcagag cacagcagag 1380
ccaacagccc caccagaaga gagcttcagg ttggggagg agaaaacaa tccctctcac 1440
aagcaggagc cgatagacaa ggaactgtat cctttaactt ccttcagatc actctttgac 1500
aacgacccct cgtcacagta aggatcggcg gccagctcaa ggaggcgctg ctgcgacgcc 1560
tcggccgacga caccgtgctg gaggagatga acctgcccgc caagtggaa gccaagatga 1620
cggcggggat cgggggcttc atcaaggtgc ggcagtagca ccagatcccc gfggagatct 1680
gcggccacaa ggccatcggc acgctgctgg tgggcccacc cccgctgaac atcatcgccc 1740
gcaacctgtc gaccacatc ggctgcaccc tgaacttccc catcagcccc atcgagacgg 1800

```

tgcccgtgaa gctgaagccg gggatggacg gccccaaggt caagcagtgg cccctgtaag 1860
aattc 1865

<210> 79
<211> 1865
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: GP2

<400> 79
gtcgacgccca ccatggggcgc cgcgcgccagc gtgctgagcg gcggcgagct ggacaagtgg 60
gagaagatcc gctctgcgccc cggcgggcaag aagaagtaca agctgaagca catcgtgtgg 120
gccacgcgcgc agctggagcgc cttcgccgtg aaccccgccc tgcctggagac cagcgaggcgc 180
tgccgcgaga tctctgggcca gctgcagccc agcctgcaga ccggcagcga ggagctgcgc 240
agcctgtaca acacgctggc caccctgtac tgcgtgcacc agcgcatcga cgtcaaggagc 300
accaagaggag cctctggagaa gatcgaggag gaggcaagaa agtccaagaa gaaggcccgag 360
caggccgcgcg ccgcgcgcgc caccggcaac agcagccagg tsagccagaa ctaccccatc 420
gtgcagaaacc tgcaggggcca gatgggtgcac caggccatca gcccccgagc cctgaacgoc 480
tgggtgaagg tgggtggagga gaaggccttc agcccccgagg tgatcccccac gttcagcgcc 540
ctgagcgagg ggcgcaacccc ccaggacctg aacacgatgt tgaacaccgt gggcgggccac 600
caggccgcgca tgcagatgct gaaggagacc atcaacgagg agggcccgcca gtgggaccgc 660
gtgcaccccc tgcacgcgcgc ccccatcgcc ccgggccaga tgcgcgagcc ccgcggcagc 720
gacatcgccg gcaccaccag caccctgcag gaggcagatc gctggatgac caacaacccc 780
cccatccccg tggggcgagat ctacaagcgg tggatcatcc tgggacctgaa caagatcgctg 840
cggatgtaca gccccaccag catcctggac atccgcgagg gccccaaagga gcccttccgc 900
gactacgtgg accgcttcta caagaccctg cgcgctgagc agggcagcca ggacgtgaag 960
aactgagtga cggagaccct gctggtgcag aacgcacaacc ccgactgcaa gaccatcctg 1020
aaggctctcg gcccccgggc caccctggag gagatgatga ccgctgcaga gggcgctgggc 1080
ggccccggcc acaaggccgc cgtgctggcc gaggcgatga gccaggtgac gaaccggcgc 1140
accatcatga tgcagcgccg caacttccgc aaccagcga agaccgtcaa gtgcttcaac 1200
tgccgcaagg agggccacac cgccaggaaac tgcgcgcgcc ccgcgaagaa gggcgtgtgg 1260
cgctgcggcc gcgaaggaca ccaaatgaaa gattgcactg agagacaggc taatttttta 1320
ggaaagatct ggccttctta caagggaagg ccagggaatt ttcttcagag cagaccagag 1380
ccaacagccc caccagaaga gagcttcagg ttgggggagg agaaaacaac tcctcttcag 1440
aagcaggagc cgatagacaa ggaactgtat cctttaactt ccttcagatc atctcttggc 1500
aacgaccccc cgtcacagta aggatcgggg ggcaactcaa ggaagcgctg ctcgatacag 1560
gagcagatga tacagtatta gaagaaatga atttgccagg aaatgggaaa ccaaaaaatga 1620
taggggggat cgggggcttc atcaaggtga ggcagtagca ccagatacct gtagaatact 1680
gtggacataa agctataggt acagatttag taggaactac acctgtcaac ataatgggaa 1740
gaaatctggt gaccacagatc ggtgcacct tgaacttccc catcagccct attgagacgg 1800
tgcccgtgaa gttgaagcgc gggatggacg gccccaaggt caagcaatgg ccatgtgaag 1860
aattc 1865

<210> 80
<211> 2305
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
FS(+).proinact.RTopt.YM

<400> 80
gcggccgcga aggaacaccaa atgaagatt gcaactgagag acaggctaat tttttaggga 60
agatctggcc ttctatacag ggaaggccag ggaattttct tcagagcaga ccagagccaa 120
cagccccacc agaagagagc ttcaggtttg gggaggagaa acaactccc tctcagaagc 180
aggagcgcat agacaaggaa ctgtatcctt taacttccct cagatacctc ttgggcaagg 240

```

acccctcgct acaataagga tcgggggggca actcaaggaa gcgctgctcg atacaggagc 300
agatgatata gtattagaag aaatgaattt gccaggaaaa tggaaaccaa aatatgatgg 360
ggggatcggg ggcttcatac aggtgaggca gtacgaccag atacctgtag aaatctgtgg 420
acataaagct ataggatcac tattagttag acctacacct gtcaacataa ttggaagaaa 480
tctgttgacc cagatcggtc gcaacctgaa ctccccatc agccctattg agacggtgcc 540
cgtgaagtgt aagccgggga tggacggccc caaggtcaag caatggccat tgaccggagg 600
gaagatcaag gccctggtgg agatctgcac cgagatggag aaggagggca agatcagcaa 660
catcgccccc gagaaacctt acaacacccc cgtgttcgcc atcaagaaga aggacagcac 720
caagtggcgc aagctggtgg acttcgcgga gctgaacaag cgcacccagg acttctggga 780
ggtgcagctg ggcaccccc accccgcggc cctgaagaag aagaagagcg tgaccgtgct 840
ggacgtgggc gacgcctact tcagcgtgcc cctggacaag gacttcgcga agtacaccgc 900
cttcacatcc cccagcatca acaacgagac ccccgccatc cgtaccagt acaacgtgct 960
ggcccggggc tggaaaggga gccccgccat ctccagagc agcatgacca agatcctgga 1020
gcccttcggc aagcagaacc ccgacatcgt gatctaccag gcccccctgt acgtggcgag 1080
cgactctgga ctcggccagc accgcaccaa gatcgaggag ctgcgccagc agtctgtcgc 1140
ctggggcttc accaccccc acaagaagca ccagaaggag ccccccttcc tgtgatggg 1200
ctacgagctg caccgcgaca agtggaccgt cgagccctac atgctgcccg agaaggacag 1260
ctggaccgtg aacgacatcc agaagctggt gggcaagctg aactgggcca gccagatcta 1320
cgccggcatc aaggtgaagc agctgtgcaa gctgctgcgc ggcaccaagg cctgaccca 1380
gggatctccc ctgaccgagg aggcgcagct ggagctggcc gagaaaccgc agatcctgaa 1440
ggaccgccgt cagcaggtgt actacgacct cagcaaggac ctggtggcgc agatccagaa 1500
cgaggggcag ggcagtgga cctaccagat ctaccaggag cccttcaaga acctgaagac 1560
ggcgaagtac gcccgcatgc gggcgcccca caccacacag gtgaagcagc tgaccgagcg 1620
cgtgcagaag tggagcaccg agagcatcgt gatctggggc agatccccca agtccaagct 1680
gcccatccag aaggagacct gggaggccgt gtggatggag tactgvcagg ccacctggat 1740
ccccgagggt gagtctgtga acaccccccc cctggtgaag ctgtggtacc cctggagaaa 1800
ggagcccatc gtggggcgcc agaccttcta cgtggacgcg gcgcaccaac gcgagaccaa 1860
gctgggcaag gcccgctacg tgaccgacgc gggcgccgca aaggtggtga gcatcgccga 1920
caccaccaac cagaagacgc agctgcagcg catccacctg gccctgcagg acagcgccct 1980
ggaggtgaac atcgtgacgc acagccagta cgccctgggc atcatccagg cccagcccca 2040
caagagcgag agcgagctgg tgagccagat catcgagcag ctgatcaaga aggaaagggt 2100
gtacctggcc tgggtgcccg cccacaaggc catcgccgcg aacgagcagc tggacaagct 2160
ggtgagcgcc ggcaccccca aggtgtggt cctgaaccgc atcgatggcg gcatcgtgat 2220
ctaccagtac atggacgacc tgtactgtgg cagcgccgcg cctaggatcg attaaaagct 2280
tcccgggctc agcaccggtg aattc 2305

```

<210> 81

<211> 2299

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

FS (+).proinact.RTopt.YMMH

<400> 81

```

gcggccgcga aggacaccaa atgaaagatt gcactgagag acaggctaat tttttaggga 60
agatctggcc ttctatacag ggaaggccag ggaattttct ctagagcaga ccagagccaa 120
agcctccacc agaagagagc ttacggtttg gggaggagaa acaactccc ttctagaagc 180
aggagcgcat agacaaggaa ctgtatcctt taacttccct cagatcactc ttgggcaagc 240
acccctcgte acaataagga tcgggggggca actcaaggaa gcgctgctcg atacagagc 300
agatgataca gtattagaag aaatgaattt gccaggaaaa tggaaaccaa aaatgatagg 360
ggggatcggg ggcttcatac aggtgaggga gtacgaccag atacctgtg aaactctggt 420
acataaagct ataggtacag tattagttag acctacacct gtcaacataa ttggaagaaa 480
tctgttgacc cagatcggtc gcaacctgaa ctccccatc agccctattg agacggtgcc 540
cgtgaagtgt aagccgggga tggacggccc caaggtcaag caatggccat tgaccagaga 600
gaagatcaag gccctggtgg agatctgcac cgagatggag aaggagggca agatcagcaa 660
catcgccccc gagaaacctt acaacacccc cgtgttcgcc atcaagaaga aggacagcac 720
caagtggcgc aagctggtgg acttcgcgga gctgaacaag cgcacccagg acttctggga 780

```

```

ggtgcagctg ggcattcccc accccgcggc cctgaagaag aagaagagcg tgaccgtgct 840
ggacgtgggg gacgcctact tcagcgtgcc cctggacaag gacttcgcga agtacaccgc 900
cttcaccatc ccagcatca acaacagagc ccccgccatc cgctaccagt acaacgtgct 960
gccccaggcg tggaaaggga gccccgccat ctccagagc agcatgacca agatcctgga 1020
gcccttcgcg aagcagaacc ccgacatcgt gatctaccag gcccccctgt acgtggggcag 1080
cgacctggag atcgccacc accgcaccaa gatcgaggag ctgcgccagc acctgctgcg 1140
ctgggggttc accaccccc acaagaagca ccagaaggag ccccccttc tgcctactga 1200
gctgcacccc gacaagtgga cgtgcagccc catcatgctg ccgagaaggg acagctggag 1260
cgtgaacgac atccagaagc tgggtgggcaa gctgaactgg gccagccaga tctacgcggg 1320
catcaagctg aagcagctgt gcaagctgct gcgcggcacc aaggccctga ccgaggtgat 1380
ccccctgacc gaggaggcgg agctggagct ggccgagaac cgcagatccc tgaaggagcc 1440
cgtgcacgag gtgtactacg accccagcaa ggacgtggtg gccagatccc agaagcaggg 1500
ccaggggcag tggacctacc agatctacca ggagcccttc aagaacctga agaccggcaa 1560
gtacgcccgc atgcgcggcg cccacaccaa cgacgtgaag cagctgaccg aggcgctgca 1620
gaaggtggag accgagagca tctgtgatct gggcaagatc cccaagtcca agctgcccat 1680
ccagaaggag acctgggagg cctggtggat ggagtactgg caggccacct ggatccccga 1740
gtgggagttc tctgaacccc cccccctggt gaagctgtgg taccagctgg agaaggagcc 1800
catcgtggcg gccagagacct tctacttgga cggcgccgcg aaccgcgaga ccaagctggg 1860
caaggcgccg tacgtgacgc accggggcgg gcagaagggt gtgagcatcg ccgacaccaa 1920
caaccagaag accgagctgc aggccatcca cctggccctg caggacagcg gcttgagggt 1980
gaacatcgct accgacagcc agtaagccct gggcatcacc caggccccag ccgacaagag 2040
cgagaagcag ctgggtgagcc agatcatcga cagcgtgact aagaaggaga aggtgtacct 2100
ggcctgggtg cccgccacca agggcatcgg cggcaacagc caggtggaca agctggtgag 2160
gcgcgggctc cgaaagggtc tgttctctga cggcatcgat ggccgcatcg gttactacca 2220
gtacatggac gacctgtacg tgggcagcgg cggccctagg atcgattaaa agcttccccg 2280
ggctagcacc ggtgaattc

```

<210> 82
 <211> 2306
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> Description of Artificial Sequence:
 FS (-).protmod.RTopt.YM

```

<400> 82
cggggccgga aggcacacaa atgaagatt gcactgagag acaggetaat ttcttcggcg 60
aggacctggc cttcctgcag ggcaaggccc gcaggttcag cagcgagcag acccgcgcca 120
acagcccacc ccgcgcggag ctgcaggtgt gggggcggga gaacaacagc ctgagcgagg 180
ccggcgcgga ccgcagggg cccgtgagct tcaacttccc ccagatcacc ctgtggcagc 240
gccccctggt gacatcagg atcggcgccc agctcaaggga ggcgtgtctc gacaccggcg 300
ccgacagacac cgtgctggag gagatgaacc tgcccggcaa gtggaaagccc aagatgatcg 360
cggggatcgg gggcttcacc aaggtgcggc agtacgacca gatccccgtg gagatctcgg 420
gccacaaggc catcgccacc gtgctggtgg gcccccacc cgtgaacatc atcggccgca 480
acctgctgac ccagatcygc tgcaccctga acttccccat cagcccccat gagacgtgct 540
cgtgaaagct gaagccgggg atggacggcc ccaaggtcaa gcagtggccc ctgaccaggg 600
agaagatcaa ggccccgtgt gagatctgca ccgagatgga gaaggagggg aagatcagca 660
agatcgcccc cgagaacccc tacacaacccc ccgtgttgc catcaagaag aaggacagca 720
ccaagtggcg caagctggtg gacttcgcgg agctgaacaa gcgcacccag gactctggtg 780
aggtgacagt ggcatcccc cacccccggc gcttgaagaa gaagaagagc gattccgtgc 840
tgagcgtggc cgacgcctac ttcagctgcg ccttggacaa ggacttcgcg aagtacaccg 900
ccttcaccat ccccagcatc aacaacagaa cccccggcat ccgctaccag tacaaagctg 960
tgccccaggg ctggaagggg agcccccgca tcttcacag cagcatgacc aagatcctgg 1020
agcccttcgg caagcagaac cccgacatcg tgacttacca ggccccctgt tacgtgggca 1080
cgacactgga gatcgccag caccgcacca agatcgagga gctgcgccag cactgctgct 1140
gctggggctt caccaccccc gacaagaagc accagaaggg gccccccttc ctgtgaggtg 1200
gctacagagt gcacccagc aagtggacgg tgacagccat catgctgccc gagaaggaca 1260
gctggacgct gaacgacatc cagaagctgg tgggcaagct gaactgggac agccagatct 1320

```

```

acgccggcat caaggtgaag cagctgtgca agctgtcgcg cggcaccacg gccctgacgg 1380
aggtgatccc cctgaccgag gaggccgagc tggagctggc cgagaccgcg gagatcctga 1440
aggagcccgct gcacgaggtg tactacgacc ccagcaagga cctggctggcc gagatccaga 1500
agcaggggcca gggccagctgg acctaccaga tctaccagga gccctcaag aacctgaaga 1560
cggccaagta cggcccgcatg cggcgcgccc acaccaacga cgtgaagcag ctgaccgagg 1620
cogtgcagaa ggtgagcacc gagagcatcg tgatctgggg caagatcccc aagtccaagc 1680
tgcccatcca gaaggagacc tgggaggctc ggtggatgga gtactggcag gccacctgga 1680
tccccagctg ggaattctgtg aacaccccccc cctgggtgaa gctgtggtag cagctggaga 1800
aggagcccat cgtggggcgcc gagaccttct acgtggacgg cgcccgcaac cggcagacca 1860
agctgggcaa gggcggtacc gtgaccgacc ggggcgggga ggaagtggtg agcatcgccc 1920
acaccacca cagagaagacc gagctgcagg ccatccacct ggcctgacg gacagcgccc 1980
tggaggtgaa catcgtgacc gacagccagt acgcccctgg catcatccag gccacggccc 2040
acaagagcga gaggcagctg gtgagccaga tcatcgagca gctgatcaag aaggagagg 2100
tgtacctggc ctgggtgccc gccaccaagg gcatcgggcg caacgagcag gtggacagg 2160
tggtagcgcc cggcatccgc aaggtgctgt tctctgaacg catcgatggc ggcactgtga 2220
ttaccagta catggcagac ctgtacgtcg gcagcggcgg ccctaggatc gattaaaaag 2280
tccccggggc tagcaccggt gaattc

```

<210> 83

<211> 2300

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

FS(-).protmod.RTopt.YMM

<400> 83

```

gcccccgga aggacacca atgaagatt gcaactgagag acaggctaat ttcttcggcg 60
aggactctggc ctctctgag ggcaaggccc cgcgagctcag cagcgagcag acccgcgcca 120
acagcccac cgcgcgcgag ctgcaggtgt gggcgcgga gaacacacgc ctgagcgagg 180
ccggcgccca gcgcagggc accgtgagct tcaacttccc ccagatcacc ctgtgaccgc 240
gccccctggt gaccatcagg atcgcgccgc agctcaagga ggcgctgctc gacacggcgg 300
ccgacgacca cgtgctggag gagatgaacc tgcccgga gttggaagccc aagatgatcg 360
gccccatcgg gggcttcac aaggtgcggc agtacgacca gatccccgtg gagatctcgg 420
gccacaaggc catcggcacc gtgctggtgg gcccccacccc cgtgaacatc atcgccgcga 480
acctgctgac ccagatcgcc tgcacctga acttccccat cagcccccat gagacggtgc 540
cogtgaagct gaagccgggg atggacggcc ccaaggtcaa gcagtggccc ctgaccgagg 600
agaagatcaa ggcctcggtg gagatctgca ccagatgga gaaggaggcg aagatcagca 660
agatcgcccc cgagaacccc tacaacaccc cctgttctgc catcaagaag aaggacagca 720
ccaagtggcg caagctgggt gaactccggc agctgaacaa gcgcacccc gaactctggg 780
aggtgcagct gggcatcccc caccgccggc gcctgaagaa gaagaagagc gtgaccgtgc 840
tggagctggg cgcgcctacc ttacagctgc cctggacaaa ggacttccgc aagtacacgc 900
cttccacct cccagcacc aacaacgaga ccccggcgat ccgctaccag tacaaactgc 960
tgccccagct ctggaagggg agcccccgca tcttccagag cagcatgacc aagatcggcg 1020
agcccttcgg caagcagaac cccgacatcg tgatctacca gggccccctg tacgtgggca 1080
gcgacctgga gatcgccag caccgcacca agatcgagga gctgcgccc cactgctgc 1140
gctggggctt caccaccccc gacaagaagc accagaagga gccccccttc ctgccatcg 1200
agctgcaccc cgacaagtgg accgtgcagg ccatcatgct gcccgagaag gacagctgga 1260
cgtgaacga catccagaag ctgggtggga agctgaactg ggcagccagc atctacggcg 1320
gcatcaaggt gaagcagctg tgcaagctgc tgcgcgccac caaggccctg accgaggtga 1380
tccccctgac cgaggaggcc gagctggagc tggccgagaa cgcgagatc ctgaaggagc 1440
cgtgcacga ggtgtactac gacccagca agaacctggt cgcgagatc cagaagcagc 1500
gccagggcca gtggacctac cagatetacc aggaacctt caagaacctg aagacggca 1560
agtacccccg ctgctggcgc gccacaccca acgacgtgaa gcagctgacc gacgocgtgc 1620
agaaggttag caccgagagc atcgtgatct ggggcaagat ccccaagttc aagctcgcca 1680
tcagaagga gacctgggag gccctggtaga ttgagtaact gcaggccacc tggatccccg 1740
agtgggagtt cgtgaacacc cccccctgg tgaagctact gtaccagctg tgagaaggagc 1800
ccatcgtggg cgcgcagacc ttctacgtgg acggcgcgcc caaccgagag accaagctgg 1860

```



```

gcaaggccgg ctacgtgacc gaccggggcc ggcagaaggt ggtgagcacc gccgacacca 1920
cccaaccagaa gaccgagctg caggccatcc acctggccct gcaggacagc ggcctggagg 1980
tgaacatcgt gaccgacagc cagtacgccc tggggcatcat ccaggccagc ccgcacaaga 2040
gcgagagcga gctgtggtag cagatcatcg agcagctgat caagaaggag aaggtgtacc 2100
tggtcctgggt gccgcgccac aagggtcatcg gcggcaacga gcaggtggag aagctggtag 2160
gcgcggccat ccgcaagggt ctgttctcga accgcatcga tggcggcacc gtgatctacc 2220
agatcatgga cgacctgtac gtgggcagcg gcggccctag gatcgattaa aagcttcccg 2280
gggttagcac cgtggaattc

```

<210> 84

<211> 2312

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

FS (-).protmod.RTopt(+)

<400> 84

```

ggggcccgga aggcacccaa atgaaagatt gacttgagag acaggctaat ttcttccgcg 60
aggacctggc ctctctcgag ggcaaggccc gcgagttcag cagcgagcag acccgcgcca 120
acagcccccac ccgcgcggag ctgcagggtg ggggcggcga gaacaacagc ctgagcgagg 180
ccggcccgca ccgaccgggc accgtgagct tcaacttccc ccagatcacc ctgtggcagc 240
gccccctggg gacccatcagg atcgcgccgc agctcaaggga ggcgctgctc gacaccggcg 300
cgcgacgacac cgtgctggag gagatgaacc tgcccggcaa gtggaaagccc aagatgatcg 360
gcgggatcgg gggcttcacc aaggtgcgcc agtacgacga gatccccgtg atcgcccgca 420
gccacaaggc catcggcacc gtgtctgttg gccccaaccc cgtgaacatc atcgcccgca 480
acctgtctgac ccagatcggc tgcaacctga acttccccat cagcccatc gagagcggtc 540
cggtgaagct gaagccgggg atggacggcc ccaagggtcaa gcagtggccc ctgaccgagg 600
agaagatcaa ggcctgggtg gagatctgca ccgagatgga gaaggaggcg aagatcagca 660
agatcgcccc cgaaaccccc tacaacaccc ccgtgttcgc catcaagaag aaggacagca 720
ccaagtggcg caagctgggt gacttccgcg agctgaacaa gcgcacccag gacttctggg 780
aggctgcagct gggcatcccc caccocgcgc gcttgaagaa gaagaagagc gtgaccgtgc 840
tgagcctggg cgaagcctac ttcagcgtgc ccttggacaa ggacttccgc aagtacaccg 900
cttccacatc cccagcacc aacaacgaga ccccgccgat ccgctaccag tacaacgtgc 960
tgccccaggc ctggaagggc agcccccgca ttctccagag cagcatgacc aagatctcgg 1020
agccctctcc caagcagaac ccgacatcg tgatctacca gtacatggag gacctgtacg 1080
tgggcagcga cctggagatc ggccagcacc gaccaagat cgaggagctg ccgaccagcc 1140
tgctgcgctg gggcttcacc acccccgaca agaagcacca gaaggagccc ccttctcgt 1200
ggatgggcta cgagctgcac ccgacaaagt ggaacctgca gcccatcatg ctgcccagca 1260
aggacagctg gacgctgaac gacatccaga agctggtggg caagctgaac tggggcagcc 1320
agactctacg cggcatcaag gtgaagcagc tgtgcaagct gctgcgcggc accaaggccc 1380
tgacggaggt gatccccctg accgagagg ccgagctgga gctggccgag aaccgacga 1440
tcttgaagga gccctgtcac gagggtgtact acgacccag caaggacctg gtggccgaga 1500
tccagaagca gggccagggc cagtggacct accagatcta ccaggagccc ttcaagaacc 1560
tgaagaccgg caagtacgac ccgatcgcg cgcccccac caacgacgtg aagcagctga 1620
ccgagcgctg gcagaagggt agcacccaga gcatcgtag ctggggcaag atccccagt 1680
tcaagctgcc catccagaag gagacctggg aggcctggtg gatggagtac tggcagcgca 1740
cctggatccc cgagtgggag ttctgtgaaca ccccccccgt ggtgaagctg ttgtaccagc 1800
tggaagaagg gcccatcgtg ggcgcccaga ccttctacgt ggaaggcgcc gccaacgcgc 1860
agaccacgct gggcaaggcc ggctactgta ccgaccgggg ccgagagaag gtggtgagca 1920
tcgcccagac caccaaaccg aagaccgagc gccagtacgc cctgggcatc atccaggccc 2040
agaccgacaa cctggcctgt gtgccgcgcc acagggcatc cgagcagctg atcaagaagg 2100
agaagtgta cctggcctgt gtgccgcgcc acagggcatc cgagcagctg gatggcggc 2160
acaagctggt gagcgccgc atccgcaagg tgcgttctc gaacggcacc gatggcgga 2220
tctgatctca ccagtagctg gacgacctgt agtggggcag cggcgccccc aggtacgatt 2280
aaaagcttcc cgggctagc accggtgaat tc

```

2312

<210> 85
 <211> 306
 <212> DNA
 <213> Human immunodeficiency virus

<400> 85
 atggagccag tagatcctag attagagccc tggaagcatc caggaagtca gcttaagact 60
 gcttgataaa attgctattg taaaaagtgt tgctttcatt gccaaagtttg ttctcataaca 120
 aaaggcttag gcatctccta tggcaggaag aagcggagac agcgacgaag agctcctcca 180
 gcagctgagg ttcatcaagt ttctctacca aagcaacccg cttccagacc ccaaggggac 240
 ccgacaggcc cgaaggaatc gaagaagaag gtggagagag agaccagagac agatccagtc 300
 cattaag 306

<210> 86
 <211> 101
 <212> PRT
 <213> Human immunodeficiency virus

<400> 86
 Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser
 1 5 10 15
 Gln Pro Lys Thr Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Phe
 20 25 30
 His Cys Gln Val Cys Phe Ile Thr Lys Gly Leu Gly Ile Ser Tyr Gly
 35 40 45
 Arg Lys Lys Arg Arg Gln Arg Arg Ala Pro Pro Asp Ser Glu Val
 50 55 60
 His Gln Val Ser Leu Pro Lys Gln Pro Ala Ser Gln Pro Gln Gly Asp
 65 70 75 80
 Pro Thr Gly Pro Lys Glu Ser Lys Lys Lys Val Glu Arg Glu Thr Glu
 85 90 95
 Thr Asp Pro Val His
 100

<210> 87
 <211> 306
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: tat.SF162.opt

<400> 87
 atggagcccg tggacccccg cctggagccc tggaagcacc ccggcagcca gcccaagacc 60
 gcctgcacca actgctactg caagaagtgc tgcttccact gccaggtgtg cttcatcacc 120
 aaagggcctgg gcatcagcta cgcccgcaag aagcgcgcgc agcgccgccc cgcccccccc 180
 gcacgcagag tgcaccaggt gagcctgccc aagcagcccc ccagccagcc ccagggcgac 240
 cccacgggcc ccaaggagag caagaagaag gtggagcgcg agaccgagac gcacccccgt 300
 cactag 306

<210> 88
 <211> 306

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

tat.cys22.SF162.opt

<400> 88

```

atggagcccg tggacccccc cctggagccc tggaaagcacc ccggcagcca gcccaagacc 60
gccggcacca actgctactg caagaagtgc tgcttcact gccaggtgtg cttcatcacc 120
aagggcctgg gcatcagcta cggccgcaag aagcgcgcgc agcgcgcgc cgcccccccc 180
gacagcgagg tgcaccaggt gagcctgccc aagcagcccg ccagccagcc ccaggggcgc 240
cccaccggcc ccaaggagag caagaagaag gtggagcgcg agaccgagac cgaccocgtg 300
cactag                                     305

```

<210> 89

<211> 168

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

tatamino.SF162.opt

<400> 89

```

atggagcccg tggacccccc cctggagccc tggaaagcacc ccggcagcca gcccaagacc 60
gccggcacca actgctactg caagaagtgc tgcttcact gccaggtgtg cttcatcacc 120
aagggcctgg gcatcagcta cggccgcaag aagcgcgcgc agcgcgcgc 168

```

<210> 90

<211> 102

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: tat cys22

SF162 protein

<400> 90

```

Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser
 1             5             10             15

Gln Pro Lys Thr Ala Gly Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe
 20             25             30

His Cys Gln Val Cys Phe Ile Thr Lys Gly Leu Gly Ile Ser Tyr Gly
 35             40             45

Arg Lys Lys Arg Arg Gln Arg Arg Arg Ala Pro Pro Asp Ser Glu Val
 50             55             60

His Gln Val Ser Leu Pro Lys Gln Pro Ala Ser Gln Pro Gln Gly Asp
 65             70             75             80

Pro Thr Gly Pro Lys Glu Ser Lys Lys Lys Val Glu Arg Glu Thr Glu
 85             90             95

Thr Asp Pro Val His Glx
 100

```